Quantization from the algebraic viewpoint

L. Abellanas and L. Martinez Alonso

Departamento de Física Teórica, Universidad Complutense de Madrid, Madrid, Spain
(Received 13 February 1976)

We use the Weyl quantization W in a general context valid for any finite-dimensional Lie algebra G, to derive an explicit formula for $(P_1, P_2) \equiv W^{-1}([W(P_1), W(P_2)])$, P_1, P_2 polynomials. In the particular case of the Heisenberg Lie algebra, this formula reduces to the familiar Moyal bracket.

1. THE VON NEUMANN QUANTIZATION

Let G be a complex Lie algebra with commutation relations $[A_i, A_j] = i \sum_k c_{ij}^k A_k$, in a given basis $\{A_j\}_1^n$ Its symmetric algebra S, isomorphic to the polynomial ring $\mathbb{C}[a_1, \ldots, a_n]$ in n commuting variables $\{a_j\}_1^n$, admits a Poisson structure[1] defined by the Poisson brackets:

$$\{P_1, P_2\} \equiv \sum_{i,j,k} c_{ij}^k a_k \frac{\partial P_1}{\partial a_i} \frac{\partial P_2}{\partial a_j}, \quad P_1, P_2 \in S.$$

On the other hand, the universal enveloping algebra[2] U of G includes all (noncommutative) polynomials in the elements of G, and has natural Lie algebra structure given by

$$[u, v] \equiv uv - vu, \quad u, v \in U.$$

Suppose a quantization rule $a_j \to A_j$, $j = 1, \ldots, n$, is fixed. Following von Neumann,[3] we are interested in finding all algebraic quantization prescriptions compatible with it, in the following sense:

Definition 1: A linear map $\rho : S \to U$, defined on S is said to be a von Neumann G-quantization if and only if

(1) $\rho(a_j) = A_j$, $\forall j$,

(2) $\rho(P^m) = (\rho(P))^m$, $\forall P \in S$, $\forall m \in \mathbb{N}$,

It turns out, however, that these requirements are incompatible. Therefore, as it will be explicitly shown below, a given Lie algebra G does not necessarily admit any such quantization. In particular it is so for the Heisenberg Lie algebra $\{P, Q, 1\}$.

It is because we propose a weaker version of von Neumann notion:

Definition 2: A linear map $\sigma : S \to U$, defined on S will be called a weak von Neumann G-quantization for G if and only if it verifies

$$\sigma\left(\left(\sum_j \xi_j a_j\right)^m\right) = \left(\sum_j \xi_j A_j\right)^m, \quad \forall \xi \equiv (\xi_1, \ldots, \xi_n) \in \mathbb{C}^n.$$

Proposition 1: For a given G there exists a unique weak von Neumann G-quantization. Furthermore, it is characterized by the symmetrization:

$$\sigma(a_{j_1} a_{j_2} \cdots a_{j_r}) = \frac{1}{r!} \sum_\pi A_{j_{\pi(1)}} A_{j_{\pi(2)}} \cdots A_{j_{\pi(r)}}, \quad (1)$$

where the summation runs over all permutations π of r objects.

Proof: It is easily verified that (1) defines a weak von Neumann G-quantization. Uniqueness follows from the fact that every polynomial $P(a_1, \ldots, a_n)$ can

be written as a linear combination of terms of the form $\left(\sum_j \xi_j a_j\right)^m$ (QED).

We are already in a position to prove that the Heisenberg Lie algebra $G_H \equiv \{A_1, A_2, A_3\}$, $[A_1, A_2] = iA_3$, $[A_1, A_3] = [A_2, A_3] = 0$, does not admit any G-quantization of the type included in Definition 1. In fact, as a simple calculation shows,

$$\sigma(a_1^2 a_2^2) = A_2^2 A_1^2 + 2iA_3 A_2 A_1 - \tfrac{1}{2}A_3^2 = (\sigma(a_1 a_2))^2 - \tfrac{1}{4}A_3^2.$$

2. THE WEYL QUANTIZATION

At this point we make some useful notation conventions:

$$\alpha \equiv (\alpha_1, \ldots, \alpha_n) \in \mathbb{N}^n, \quad \alpha! \equiv \alpha_1! \alpha_2! \cdots \alpha_n!, \quad |\alpha| \equiv \sum_j \alpha_j,$$

$$\xi A \equiv \sum_j \xi_j A_j, \quad \xi^\alpha \equiv \xi_1^{\alpha_1} \xi_2^{\alpha_2} \cdots \xi_n^{\alpha_n}, \quad \forall \xi \in \mathbb{C}^n,$$

$$\frac{\partial^{|\alpha|}}{\partial \xi^\alpha} \equiv \frac{\partial^{|\alpha|}}{\partial \xi_1^{\alpha_1} \partial \xi_2^{\alpha_2} \cdots \partial \xi_n^{\alpha_n}}.$$

As it is well known[2] $U = \oplus_m U^m$, where U^m stands for the linear subspace of U generated by the elements $\sigma(a^\alpha)$, $|\alpha| = m$. Let us consider the set U^* of all formal series in the $\sigma(a^\alpha)$. As a typical example we quote

$$e^{i\xi A} \equiv \sum_{m=0}^\infty \frac{i^m}{m!} (\xi A)^m = \sum_\alpha \frac{i^{|\alpha|}}{\alpha!} \xi^\alpha \sigma(a^\alpha), \quad \xi \in \mathbb{R}^n. \quad (2)$$

The underlying idea in what follows is to use the right-hand side expression in order to freely manipulate the coefficients without interference of the noncommutative part $\sigma(a^\alpha)$. Thus, given a set of complex-valued functions $F_\alpha : \mathbb{R}^n \to \mathbb{C}$, $\alpha \in \mathbb{N}^n$, we define an associated map:

$$F : \mathbb{R}^n \to U^*, \quad F(\xi) \equiv \sum_\alpha F_\alpha(\xi) \sigma(a^\alpha),$$

which assigns to each point $\xi \in \mathbb{R}^n$ a formal series in U^*. Furthermore, if our F_α are good enough, we can apply tempered distributions T to get

$$\langle T, F \rangle \equiv \sum_\alpha \langle T, F_\alpha \rangle \sigma(a^\alpha). \quad (3)$$

This produces a formal series with complex coefficients. It is in this context that Weyl quantization[4] makes sense:

Definition 3: $W : S \to U$, $W(P) \equiv (2\pi)^{-n/2} \langle \hat{P}, e^{i\xi A} \rangle$, $P \in S$, where $\hat{P}(\xi) \equiv (2\pi)^{-n/2} \int P(a) e^{-i\xi a} da$ is the Fourier transform of P. But since P was a polynomial, $\hat{P}$ is a finite order distribution with point support $\xi = 0$. Therefore, $W(P) = (2\pi)^{-n/2} \sum_\alpha (i^{|\alpha|}/\alpha!) \langle \hat{P}, \xi^\alpha \rangle \sigma(a^\alpha) \in U$ [observe that $W(P)$ is nothing but a polynomial in the A_j].

Proposition 2: $W(P) = \sigma(P), \quad \forall P \in S.$

Proof: From the familiar properties of Fourier transform,

$$\hat{a}^{\alpha}(\xi) = (2\pi)^{n/2} i^{|\alpha|} \frac{\partial^{|\alpha|}}{\partial \xi^{\alpha}} \delta(\xi),$$

we conclude that

$$W(a^{\alpha}) = \sum_{\beta} i^{|\alpha|+|\beta|} \frac{(-1)^{|\alpha|}}{\beta!} \left\langle \delta(\xi), \frac{\partial^{|\alpha|}}{\partial \xi^{\alpha}} \xi^{\beta} \right\rangle \sigma(a^{\beta}) = \sigma(a^{\alpha}).$$

Hence $W = \sigma$ on S, by linearity (QED).

3. CONVOLUTION VERSUS QUANTIZATION

Since $\sigma: S \to U$ is a bijective map,[2] the element

$$P_1 * P_2 \equiv \sigma^{-1}(\sigma(P_1)\sigma(P_2)), \quad P_2, P_2 \in S, \tag{4}$$

is well defined in S. The formal analogy between (4) and the convolution product $f * g = \mathcal{J}^{-1}(\mathcal{J}f \cdot \mathcal{J}g)$ in Fourier theory is obvious, and as a matter of fact it will be made explicit in what follows.

Proposition 3:

$$(P_1 * P_2)(a) = e^{ia\tau(i\nabla', i\nabla'')} P_1(a') P_2(a'') \big|_{a'=a''=a}$$

where $\nabla' \equiv \partial/\partial a'$, $\nabla'' \equiv \partial/\partial a''$, and $\tau(\xi, \eta) \equiv \lambda(\xi, \eta) - \xi - \eta$ denotes the nonlinear part of the exponent in the Baker—Hausdorff—Campbell[5] formula: $e^{i\xi A} e^{i\eta A} = e^{i\lambda(\xi, \eta)A}$.

Remark: Since P_1, P_2 are polynomials, the right-hand side in the previous formula has a finite number of terms, so it is also a polynomial.

Proof:

$$\sigma(P_1)\sigma(P_2) = W(P_1) W(P_2) = (2\pi)^{-n} \langle \hat{P}_1(\xi) \hat{P}_2(\eta), e^{i\lambda(\xi,\eta)A} \rangle$$

$$= (2\pi)^{-n} \sum_{\alpha} \frac{i^{|\alpha|}}{\alpha!} \langle \hat{P}_1(\xi) \hat{P}_2(\eta), \lambda^{\alpha}(\xi, \eta) \rangle \sigma(a^{\alpha}).$$

By expressing $\lambda^{\alpha}(\xi, \eta) \equiv \lambda_1^{\alpha_1}(\xi, \eta) \lambda_2^{\alpha_2}(\xi, \eta) \cdots \lambda_n^{\alpha_n}(\xi, \eta)$ as a (finite) Taylor series relative to the point $\xi + \eta$, we obtain

$$\lambda^{\alpha}(\xi, \eta) = \sum_{\beta} \frac{1}{\beta!} \left(\frac{\partial^{|\beta|} \lambda^{\alpha}}{\partial \lambda^{\beta}} \bigg|_{\lambda = \xi + \eta} \right) \tau^{\beta}(\xi, \eta).$$

Thus we can write

$$\sigma(P_1)\sigma(P_2) = (2\pi)^{-n/2} \sum_{\alpha, \beta} \frac{i^{|\alpha|}}{\alpha! \beta!} D_{\alpha\beta} \sigma(a^{\alpha}), \tag{5}$$

where

$$D_{\alpha\beta} \equiv (2\pi)^{-n/2} \langle \tau^{\beta}(\xi, \eta) \hat{P}_1(\xi) \hat{P}_2(\eta), L_{\alpha\beta}(\xi + \eta) \rangle,$$

$$L_{\alpha\beta}(x) \equiv \frac{\partial^{|\beta|} \lambda^{\alpha}}{\partial \lambda^{\beta}} \bigg|_{\lambda = x}.$$

In order to calculate $D_{\alpha\beta}$, we only need to recall the analycity of τ^{β} at the origin,[5] let us say

$$\tau^{\beta}(\xi, \eta) = \sum_{\alpha', \alpha''} t_{\alpha'\alpha''} \xi^{\alpha'} \eta^{\alpha''}$$

and some well-known properties of the Fourier transform:

$$D_{\alpha\beta} = (2\pi)^{-n/2}$$

$$\times \sum_{\alpha', \alpha''} t_{\alpha'\alpha''} \langle (\widehat{(i\nabla')^{\alpha'} P_1(a')})(\xi) (\widehat{(i\nabla'')^{\alpha''} P_2(a'')})(\eta), L_{\alpha\beta}(\xi + \eta) \rangle$$

$$= \sum_{\alpha', \alpha''} t_{\alpha'\alpha''} \langle (\widehat{(i\nabla')^{\alpha'} P_1(a')}) * (\widehat{(i\nabla'')^{\alpha''} P_2(a'')}), L_{\alpha\beta} \rangle$$

$$= \sum_{\alpha', \alpha''} \langle (t_{\alpha'\alpha''} (i\nabla')^{\alpha'} P_1(a')(i\nabla'')^{\alpha''} P_2(a'') \big|_{a'=a''=a})^{\hat{}}, L_{\alpha\beta} \rangle$$

$$= \langle (\tau^{\beta}(i\nabla', i\nabla'') P_1(a') P_2(a'') \big|_{a'=a''=a})^{\hat{}}(\xi), L_{\alpha\beta}(\xi) \rangle$$

$$= i^{|\beta|} \langle (a^{\beta} \tau^{\beta}(i\nabla', i\nabla'') P_1(a') P_2(a'') \big|_{a'=a''=a})^{\hat{}}(\xi), \xi^{\alpha} \rangle.$$

By substituting in (5) we get

$$\sigma(P_1)\sigma(P_2)$$
$$= (2\pi)^{-n/2} \sum_{\alpha} \frac{i^{|\alpha|}}{\alpha!}$$

$$\times \langle (e^{ia\tau(i\nabla', i\nabla'')} P_1(a') P_2(a'') \big|_{a'=a''=a})^{\hat{}}(\xi), \xi^{\alpha} \rangle \sigma(a^{\alpha})$$

$$= \langle (e^{ia\tau(i\nabla', i\nabla'')} P_1(a') P_2(a'') \big|_{a'=a''=a})^{\hat{}}(\xi), e^{i\xi A} \rangle$$

$$= \sigma(e^{ia\tau(i\nabla', i\nabla'')} P_1(a') P_2(a'') \big|_{a'=a''=a}).$$

The rest of the proof is trivial (QED).

We will now define a new Lie structure on the symmetric algebra S as follows. If P_1, P_2 are elements in S, also $(P_1, P_2) \equiv P_1 * P_2 - P_2 * P_1 \in S$, and has the following properties:

(1) $(P_1, P_2) = \sigma^{-1}([\sigma(P_1), \sigma(P_2)])$,

(2) $(P_1, P_2) = -(P_2, P_1)$,

(3) $(P_1, \lambda P_2 + \mu P_3) = \lambda(P_1, P_2) + \mu(P_1, P_3)$,

(4) $(P_1, (P_2, P_3)) + (P_2, (P_3, P_1)) + (P_3, (P_1, P_2)) = 0$.

Finally, by making use of the Proposition 3, we have the explicit formula

$$(P_1, P_2) = (e^{ia\tau(i\nabla', i\nabla'')} - e^{ia\tau(i\nabla'', i\nabla')}) P_1(a') P_2(a'') \big|_{a'=a''=a}. \tag{6}$$

4. THE CANONICAL EXAMPLE

Let G_H be the Heisenberg Lie algebra. In this particular case we find:

$$\lambda(\xi, \eta) = (\xi_1 + \eta_1, \xi_2 + \eta_2, \xi_3 + \eta_3 + \tfrac{1}{2}(\xi_2 \eta_1 - \xi_1 \eta_2))$$

and thus

$$\tau(\xi, \eta) = (0, 0, \tfrac{1}{2}(\xi_2 \eta_1 - \xi_1 \eta_2)).$$

The formula (6) gives us the result

$$(P_1 P_2) = \left(\exp \left[i \frac{a_3}{2} \left(\frac{\partial^2}{\partial a_1' \partial a_2''} - \frac{\partial^2}{\partial a_2' \partial a_1''} \right) \right] \right.$$

$$\left. - \exp \left[i \frac{a_3}{2} \left(\frac{\partial^2}{\partial a_1'' \partial a_2'} - \frac{\partial^2}{\partial a_2'' \partial a'} \right) \right] \right) P_1(a') P_2(a'') \bigg|_{a'=a''=a}$$

$$= 2i \left(\sin \frac{a_3}{2} \left(\frac{\partial}{\partial a_1'} \frac{\partial}{\partial a_2''} - \frac{\partial}{\partial a_1''} \frac{\partial}{\partial a_2'} \right) \right) P_1(a') P_2(a'') \bigg|_{a'=a''=a}.$$

In quantum mechanics, where $a_1 = q$, $a_2 = p$, $a_3 = \hbar$, this reads as follows:

1364   J. Math. Phys., Vol. 17, No. 8, August 1976

L. Abellanas and L. Martinez Alonso   1364

$$(P_1, P_2)(q, p)$$

$$= 2i\left(\sin\frac{\hbar}{2}\left(\frac{\partial}{\partial q'}\frac{\partial}{\partial p''} - \frac{\partial}{\partial q''}\frac{\partial}{\partial p'}\right)\right)$$

$$\times P_1(q', p')P_2(q'', p'')\Big|_{\substack{q'=q''=q \\ p'=p''=p}}.$$

This is identical with the well-known expression obtained by Moyal,[6] which can also be written in the form

$$(P_1, P_2)(q, p)$$

$$= 2i\sum_{n=0}^{\infty}\sum_{k=0}^{2n+1}\left(\frac{\hbar}{2}\right)^{2n+1}\frac{(-1)^{n+k}}{k!(2n-k+1)!}\frac{\partial^{2n+1}P_1}{\partial p^k\partial q^{2n+1-k}}\frac{\partial^{2n+1}P_2}{\partial q^k\partial p^{2n+1-k}}.$$

One easily sees that if one (or both) of the polynomials P_1, P_2 have degrees < 3, then

$$(P_1, P_2)(q, p) = i\hbar\left(\frac{\partial P_1}{\partial q}\frac{\partial P_2}{\partial p} - \frac{\partial P_1}{\partial p}\frac{\partial P_2}{\partial q}\right) = i\hbar\{P_1, P_2\},$$

where $\{P_1, P_2\}$ is the familiar Poisson bracket.

[1]M. Vergne, Bull. Soc. Math. France 100, 301 (1972).
[2]J. Dixmier, Algèbres enveloppantes (Gauthiers-Villars, Paris, 1974).
[3]J. von Neumann, Mthematische Begründung der Quantenmechanik (Springer-Verlag, Berlin, 1932).
[4]H. Weyl Gruppentheorie und Quantenmechanik (Hirzel-Verlag, Léipzig, 1928).
[5]N. Jacobson, Lie algebras (Interscience, New York, 1962).
[6]J.E. Moyal, Proc. Cambridge Phil. Soc. 45, 99 (1949).

Cross sections for screened potentials

Mark D. Semon and John R. Taylor

Department of Physics and Astrophysics, University of Colorado, Boulder, Colorado 80309
(Received 8 March 1976)

We consider scattering by screened Coulomb and screened short-range potentials. We prove, in Born approximation, that the cross section for a sharply cutoff Coulomb potential does not converge to the Rutherford cross section as the screening radius $\rho \to \infty$. On the other hand, for certain "smooth" screening functions we show that the screened Coulomb cross section does approach the Rutherford cross section as $\rho \to \infty$. In the case of short-range potentials, we prove (exactly, not just in Born approximation) that the screened cross section approaches the unscreened cross section as $\rho \to \infty$, whether the screening function is smooth or not.

1. INTRODUCTION

In this paper we discuss scattering cross sections for screened potentials and examine the limits of these cross sections as the screening radius ρ tends to infinity. The main interest in screened potentials derives from the well-known difficulties in the scattering theory of the Coulomb, and other long-range, potentials. While none of the standard methods of time dependent scattering theory apply to the Coulomb potential

$$V(r) = \gamma/r, \tag{1.1}$$

the same methods *do* apply to a screened Coulomb potential

$$V_\rho(r) = (\gamma/r)\alpha_\rho(r), \tag{1.2}$$

where $\alpha_\rho(r)$ is some function which goes to zero with r in a distance of order ρ. It is natural to hope that the differential cross section computed for the screened potential V_ρ would converge to a limit as $\rho \to \infty$ and that this limit would be the differential cross section for the pure Coulomb potential (1.1).

The idea of using screened Coulomb potentials is, of course, an old one. It dates back at least to the 1928 paper of Gordon,[1] and is described in several text books.[2,3] Nonetheless, it is only in the last few years that it has been carefully studied.[4-9]

In two recent papers[8,9] we have established some results for a wide class of screening functions $\alpha_\rho(r)$. We considered functions $\alpha_\rho(r)$ that satisfy

$\alpha_\rho(0) = 1$,

$\alpha_\rho(r) \to 0$ monotonically like $O(r^{-2-\epsilon})$
as $\rho \to \infty$ (r fixed), $\tag{1.3}$

$\alpha_\rho(r) \to 1$ as $r \to \infty$ (ρ fixed).

We examined the scattering probability

$$W_\rho(d\Omega \leftarrow \phi)$$

that a particle incident on V_ρ, with in state given by the momentum-space wavefunction $\phi(\mathbf{p})$, be scattered into a small solid angle $d\Omega$. Provided $\phi(\mathbf{p})$ is infinitely differentiable[10] and $d\Omega$ does not include the forward direction,[12] we were able to show that

$$\lim_{\rho \to \infty} W_\rho(d\Omega \leftarrow \phi) = W_{\text{Coul}}(d\Omega \leftarrow \phi) \tag{1.4}$$

where $W_{\text{Coul}}(d\Omega \leftarrow \phi)$ denotes the probability for scattering of ϕ into $d\Omega$ computed using the conventional Coulomb amplitude

$$f_{\text{Coul}} = -\gamma \exp\{2i \arg[\Gamma(1 + i\gamma)]\}$$
$$\times \exp[-i\gamma \ln \sin^2(\theta/2)]/2 \sin^2(\theta/2). \tag{1.5}$$

The result (1.4) goes a long way toward justifying the ideas of the first paragraph. Unfortunately, it does not itself quite prove that the cross section of the screened potential V_ρ approaches the Coulomb cross section as $\rho \to \infty$. In (1.4), $W_\rho(d\Omega \leftarrow \phi)$ is the probability of scattering off V_ρ for a *fixed* wavepacket ϕ. Now the definition of the cross section involves a large number of collisions with many different wavepackets. Specifically, if ϕ_b denotes a wavepacket obtained from ϕ by rigid displacement through an "impact parameter" $\mathbf{b}$, then the cross section for scattering into $d\Omega$ is defined as[13]

$$\sigma_\rho(d\Omega \leftarrow \phi) = \int d^2b\, W_\rho(d\Omega \leftarrow \phi_\mathbf{b}), \tag{1.6}$$

where the integral over impact parameters $\mathbf{b}$ runs over the plane perpendicular to the incident direction. Similarly,

$$\sigma_{\text{Coul}}(d\Omega \leftarrow \phi) = \int d^2b\, W_{\text{Coul}}(d\Omega \leftarrow \phi). \tag{1.7}$$

The result we would like to prove is that the screened cross section σ_ρ approaches σ_{Coul} as $\rho \to \infty$,

$$\lim_{\rho \to \infty} \sigma_\rho(d\Omega \leftarrow \phi) = \sigma_{\text{Coul}}(d\Omega \leftarrow \phi). \tag{1.8}$$

Clearly this follows from (1.4) *provided* we can interchange the order of the two limiting processes of integrating over $\mathbf{b}$ in (1.6) and letting $\rho \to \infty$.

Unfortunately, we have been unable to prove that the necessary interchange of limits is justified; and we have, in fact, found fairly compelling evidence that it is *not* justified for certain screening functions $\alpha_\rho(r)$. In Sec. 2 we examine the sharply cut-off Coulomb potential

$$V_\rho(r) = \gamma/r, \quad r \leq \rho,$$
$$= 0, \quad r > \rho,$$

that is, we take $\alpha_\rho(r) = \theta(\rho - r)$ [which obviously satisfies the conditions (1.3)]. We show that the cross section σ_ρ *in Born approximation* does *not* converge to σ_{Coul} as $\rho \to \infty$. In fact

$$\lim_{\rho \to \infty} \sigma_\rho(d\Omega \leftarrow \phi) = \tfrac{3}{2}\sigma_{\text{Coul}}(d\Omega \leftarrow \phi). \tag{1.9}$$

We shall show that the additional scattering off V_ρ is caused by the discontinuity in the potential at $r = \rho$.

However large we make ρ, this discontinuity causes an appreciable scattering of those packets with impact parameters **b** of the order of ρ. This scattering off the discontinuity contributes precisely the unwanted $\frac{1}{2}\sigma_{Coul}$ in (1.9).

It is perhaps not particularly surprising that a sharp screening of the Coulomb potential leads to difficulties, and one might well guess that the desired result (1.8) would hold only when the screening function $\alpha_\rho(r)$ is reasonably smooth. We have been unable to prove this result. However, we do show that the result (1.8) holds *in Born approximation*, provided the screening function $\alpha_\rho(r)$ has a derivative whose total variation tends to zero as $\rho \to \infty$. This condition excludes the sharp cutoff [for which $\alpha_\rho'(r) = -\delta(\rho - r)$] but includes any "reasonable," smooth screening function such as $\exp(-r/\rho)$ or $\exp(-r^2/\rho^2)$.

In view of these difficulties with screened Coulomb potentials it is interesting to consider some of the same questions for screened short-range potentials. In particular, if V is a short-range potential [which we define by the condition $V = O(r^{-3-\epsilon})$ as $r \to \infty$] and if

$$V_\rho(r) = V(r)\alpha_\rho(r),$$

then we can ask whether the cross section σ_ρ for V_ρ approaches that for V as $\rho \to \infty$.

Dollard[14] has shown that the scattering operator S_ρ for V_ρ converges strongly to the scattering operator S for V as $\rho \to \infty$.[15] From this one can deduce that the scattering probabilities $W_\rho(d\Omega \leftarrow \phi)$ and $W(d\Omega \leftarrow \phi)$ are equal in the limit that $\rho \to \infty$. But, just as in the Coulomb case, this does not settle the corresponding question for the cross sections. In Sec. 3, we consider a class of screening functions that includes all those of (1.3). (Note that this *includes* sharp cutoffs.) We show that, as $\rho \to \infty$, the scattering probability $W_\rho(d\Omega \leftarrow \phi)$, the scattering amplitude f_ρ, and the cross section $\sigma_\rho(d\Omega \leftarrow \phi)$ for the screened potential V_ρ all converge to the corresponding quantities, W, f, and σ, for the unscreened V.

2. SCREENED COULOMB SCATTERING IN BORN APPROXIMATION

In this section we consider the cross sections for screened Coulomb potentials, using amplitudes computed in Born approximation. We show that for the sharply cutoff Coulomb potential

$$V_\rho(r) = (\gamma/r)\theta(\rho - r) \qquad (2.1)$$

the cross section σ_ρ does *not* converge to σ_{Coul} as $\rho \to \infty$, the difference being accounted for by particles that are scattered by the discontinuity at $r = \rho$. On the other hand, we show that for *smooth* screening functions (which we shall define as functions with derivatives whose total variation goes to zero as $\rho \to \infty$) the screened cross section does converge to σ_{Coul} as $\rho \to \infty$.

It would, of course, be better to prove these results for the *exact* cross sections (rather than those computed in Born approximation), and it is unfortunate that we have so far failed to do this. Nevertheless, the Born results are presumably reliable for potentials that are sufficiently weak. In particular, the result that the

cross section σ_ρ for the sharp cutoff does not converge to σ_{Coul} in Born approximation suggests that the same must surely be true in the exact case as well.

For any given target potential the cross section for scattering a packet of shape ϕ into a solid angle $d\Omega$ is defined as the integral over impact parameters **b**

$$\sigma(d\Omega \leftarrow \phi) = \int d^2b \, W(d\Omega \leftarrow \phi_b), \qquad (2.2)$$

where $W(d\Omega \leftarrow \phi_b)$ is the probability that the incident packet ϕ_b, with impact parameter **b**, be scattered into $d\Omega$. It has been shown[16] that this integral is certainly convergent if (i) the potential is $O(r^{-3-\epsilon})$ as $r \to \infty$ (this is the case with all of our screened potentials), (ii) the packet $\phi(\mathbf{p})$ is infinitely differentiable with compact support contained in the forward half-space $p_z > 0$ (we take the incident mean momentum $\mathbf{p}_0$ to lie along the z axis), (iii) the direction of observation, defined by $d\Omega$, does not overlap the forward cone, defined by $\phi(\mathbf{p})$. We assume that these conditions are satisfied.[17]

The cross section defined by (2.2) can be written in a more familiar form: The probability W is written in terms of the scattered wavefunction, and the latter is written in terms of the amplitude f and incident packet ϕ. Standard manipulations (Ref. 3, pp. 49—51) then give the result

$$\sigma(d\Omega \leftarrow \phi) = d\Omega \int d^3p(p/p_z)|f(p\mathbf{u} \leftarrow \mathbf{p})\phi(\mathbf{p})|^2, \qquad (2.3)$$

where **u** is a unit vector in the direction of observation.

For normal short range potentials under ordinary conditions, the factors p/p_z and $f(p\mathbf{u} \leftarrow \mathbf{p})$ in (2.3) are essentially constant in the region where $\phi(\mathbf{p})$ is nonzero. Thus both factors can be taken outside the integral, with **p** replaced by $\mathbf{p}_0$. The remaining integral is the normalization integral for ϕ and the cross section reduces to the familiar $d\Omega|f(p_0\mathbf{u} \leftarrow \mathbf{p}_0)|^2$. In our case, we shall be using (2.3) with the screened amplitude f_ρ, and we shall need to be more careful since f_ρ may oscillate very rapidly when the screening radius ρ is large. However, we may remark here that in some cases (for example, in the Born approximation for a smoothly screened Coulomb potential) it can happen that, as $\rho \to \infty$, f_ρ approaches f_{Coul} *uniformly* for all **p** in the region of integration in (2.3). In these cases, it immediately follows from (2.3) that σ_ρ approaches σ_{Coul} as $\rho \to \infty$.

The sharp cutoff: We consider first the sharply cutoff Coulomb potential (2.1) for which the amplitude (in Born approximation) is known to be

$$f_\rho(\mathbf{p}' \leftarrow \mathbf{p}) = -(2m\gamma/q^2)(1 - \cos q\rho), \qquad (2.4)$$

where q is the magnitude of the momentum transfer $\mathbf{q} = \mathbf{p}' - \mathbf{p}$. Since the Coulomb amplitude (in Born approximation) is

$$f_{Coul}(\mathbf{p}' \leftarrow \mathbf{p}) = -2m\gamma/q^2, \qquad (2.5)$$

we can write

$$f_\rho = f_{Coul}(1 - \cos q\rho), \qquad (2.6)$$

displaying the well-known fact that the amplitude f_ρ for the sharply cutoff Coulomb potential is equal to the Coulomb amplitude *plus* a term proportional to $\cos q\rho$, which oscillates more and more rapidly as $\rho \to \infty$. As

1367   J. Math. Phys., Vol. 17, No. 8, August 1976

M.D. Semon and J.R. Taylor   1367

emphasized in Ref. 8, if we consider f_ρ as a distribution, then this second term goes to zero as $\rho \to \infty$. However, to compute the cross section (2.3), we must examine

$$|f_\rho(\mathbf{p}' \gets \mathbf{p})|^2 = |f_{\text{Coul}}(\mathbf{p}' \gets \mathbf{p})|^2 (1 - 2\cos q\rho + \cos^2 q\rho).$$

$$(2.7)$$

When this is inserted into (2.3), we obtain three terms. The first term alone gives exactly the Coulomb cross section. The second contains the factor $\cos q\rho$ in its integrand; this factor oscillates more and more rapidly as ρ grows, and, by the Riemann—Lebesque lemma, the integral goes to zero as $\rho \to \infty$. The third term also contains an oscillating factor, $\cos^2 q\rho$, in its integrand; unfortunately, $\cos^2 q\rho$ oscillates about the average value $\frac{1}{2}$. Thus as $\rho \to \infty$ the third integral contributes $\frac{1}{2}\sigma_{\text{Coul}}$, and the complete cross section therefore approaches

$$\lim \sigma_\rho(d\Omega \gets \phi) = \tfrac{3}{2}\sigma_{\text{Coul}}(d\Omega \gets \phi). \qquad (2.8)$$

That is, the cross section for the sharply cut-off Coulomb potential does not approach that for the pure Coulomb as $\rho \to \infty$. We shall return to examine the origin of the extra $\frac{1}{2}\sigma_{\text{Coul}}$ on the right of (2.8). First we consider the cross section for a smoothly screened Coulomb potential.

Smooth screening functions: Let us now consider a smoothly screened Coulomb potential

$$V_\rho(r) = (\gamma/r)\alpha_\rho(r),$$

where $\alpha_\rho(r)$ satisfies the three conditions (1.3) and in addition is *differentiable* with a derivative $\alpha'_\rho(r)$ which is of bounded variation on $[0, \infty)$ and whose total variation tends to zero as $\rho \to \infty$.[18] As mentioned in the Introduction, these conditions exclude the sharp cutoff $\alpha_\rho(r) = \theta(\rho - r)$ but admit, for example, an exponential screening $\alpha_\rho(r) = \exp(-r/\rho)$ or a Gaussian screening $\exp(-r^2/\rho^2)$.

The amplitude for V_ρ (in Born approximation) is

$$f_\rho(\mathbf{p}' \gets \mathbf{p}) = -(2m\gamma/q)\int_0^\infty dr \sin qr\, \alpha_\rho(r)$$

Since $\alpha_\rho(r)$ is differentiable, we can integrate by parts to give

$$f_\rho(\mathbf{p}' \gets \mathbf{p}) = -(2m\gamma/q^2)\left[1 + \int_0^\infty dr \cos qr\, \alpha'_\rho(r)\right]$$

$$= f_{\text{Coul}} + \delta f,$$

where

$$\delta f(\mathbf{p}' \gets \mathbf{p}) = -(2m\gamma/q^2)\int_0^\infty dr \cos qr\, \alpha'_\rho(r).$$

Since $\alpha'_\rho(r)$ is a function of bounded variation, it can be written as the difference of two positive, bounded, monotonically decreasing functions $\mu_\rho(r)$ and $\nu_\rho(r)$. The expression for δf can then be bounded using the second mean value theorem as follows:

$$|\delta f(\mathbf{p}' \gets \mathbf{p})| \le (2m|\gamma|/q^2)[\mu_\rho(0) + \nu_\rho(0)]$$

$$\times \sup_R \left|\int_0^R dr \cos qr\right|$$

$$\le (2m|\gamma|/q^3)[\mu_\rho(0) + \nu_\rho(0)]. \qquad (2.9)$$

Under the stated conditions both terms in the braces go to zero as $\rho \to \infty$. Therefore, the right-hand side of (2.9) tends to zero, uniformly for all $\mathbf{p}$ and $\mathbf{p}'$ in any compact regions that do not overlap (i.e., such that

$\mathbf{p}' \ne \mathbf{p}$). As we have already mentioned in connection with (2.3) this immediately guarantees the desired result, that

$$\lim_{\rho \to \infty} \sigma_\rho(d\Omega \gets \phi) = \sigma_{\text{Coul}}(d\Omega \gets \phi)$$

Behavior of the scattered wave from a sharp cutoff: To understand better why the cross section σ_ρ for a sharply cut-off Coulomb potential does not approach σ_{Coul} as $\rho \to \infty$, let us examine the scattered wave $\psi = (S - 1)\phi$ corresponding to an incident wave ϕ_b with impact parameter $\mathbf{b}$:

$$\psi_{\rho,b}(\mathbf{p}) = (i/2\pi m)\int d^3p'\, \delta(E - E')f_\rho(\mathbf{p} \gets \mathbf{p}')\exp(-i\mathbf{b}\cdot\mathbf{p}')\phi(\mathbf{p}').$$

$$(2.10)$$

Note that this depends on ρ, the radius at which the potential is cutoff, and $\mathbf{b}$, the impact parameter of the incident wavepacket. If we insert the expression (2.6) for the amplitude f_ρ, then we can write

$$\psi_{\rho,b}(\mathbf{p}) = \psi_b^{\text{Coul}}(\mathbf{p}) + \delta\psi_{\rho,b}(\mathbf{p}). \qquad (2.11)$$

The first term here is exactly the scattered wave one would compute using the Coulomb amplitude (in Born approximation) and is independent of the screening radius ρ. The second term,

$$\delta\psi_{\rho,b}(\mathbf{p}) = (i\gamma/\pi)\int d^3p'\, \delta(E - E')\,q^{-2}\cos q\rho\exp(-i\mathbf{b}\cdot\mathbf{p}')\phi(\mathbf{p}'),$$

$$(2.12)$$

is, of course, the source of our difficulties.

From the scattered wave (2.11) we can compute the scattering probabilities $W_\rho(d\Omega \gets \phi_b)$, and using the latter we can compute the cross section (2.2)

$$\sigma_\rho(d\Omega \gets \phi) = d\Omega \int d^2b \int_0^\infty p^2\,dp$$

$$\times \{|\psi_b^{\text{Coul}}|^2 + 2\,\text{Re}\,\psi_b^{\text{Coul}}\delta\psi_{\rho,b}^* + |\delta\psi_{\rho,b}|^2\}. \quad (2.13)$$

It is easily checked that the Coulomb scattered wave ψ_b^{Coul} falls off faster than any inverse power of the impact parameter $\mathbf{b}$. Thus the first term in (2.13) is convergent, yields precisely the Coulomb cross section, and has its main contribution from small impact parameters $\mathbf{b}$.

To discuss the second and third terms in the cross section (2.13), we must examine the extra piece $\delta\psi$ of the scattered wave, as given by (2.12). If we take any fixed cutoff radius ρ, then $\delta\psi_{\rho,b}$ goes rapidly to zero as $b \to \infty$, because of the oscillatory factor $\exp(-i\mathbf{b}\cdot\mathbf{p}')$ in the integrand. This guarantees that, for fixed ρ, the cross section (2.13) is finite, just as we would expect. However, as we make the screening radius ρ larger the factor $\cos q\rho$ in (2.12) also oscillates rapidly. Thus, however large we make ρ, we can expect that for certain impact parameters $\mathbf{b}$, with $b \approx \rho$, the integrand will not oscillate at all and the scattered wave $\delta\psi_{\rho,b}$ will be appreciable. This suggests that, however large ρ, the extra term $\delta\psi_{\rho,b}$ may contribute to the cross section, and that, the larger ρ, the larger the impact parameters $\mathbf{b}$ at which this contribution will occur.

To make these conclusions precise, one can analyze the integral (2.12) by the method of stationary phases. The conclusions of such an analysis can be briefly stated as follows[11]: The integral (2.12) runs over the small neighborhood of the incident mean momentum $\mathbf{p}_0$
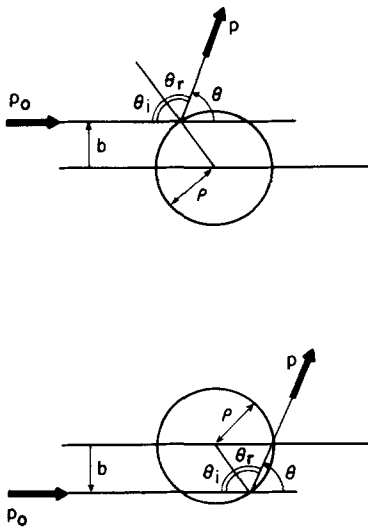
FIG. 1. The two values of the impact parameter b for which there is appreciable scattering off the discontinuity at $r = \rho$ in the cutoff potential. Note that in both cases $\theta_i = \theta_r$, and the scattering is classical specular reflection off the discontinuity.

where $\phi(\mathbf{p})$ is nonzero, and for most values of b the integrand has no stationary point inside the region of integration. As $\rho \to \infty$ these values of b make no contribution to the cross section. On the other hand, for a certain small range of impact parameters b there is a stationary point of the integrand inside the region of integration, and when b is in this small range, $\delta\psi_{\rho,b}$ is *not* negligible however large we make ρ. The values of b for which this occurs can be found most easily by evaluating $\delta\psi$ in Cartesian coordinates with $\mathbf{p}_0$ as z axis. The result is that there is appreciable scattering when b is in the immediate neighborhood of either of the two points such that

b is coplanar with $\mathbf{p}$ and $\mathbf{p}_0$,

$$b = \rho \cos\theta/2, \tag{2.14}$$

where θ is the scattering angle (the angle between $\mathbf{p}$ and $\mathbf{p}_0$). These two values of b are illustrated in Fig. 1. Some simple geometry shows that the conditions (2.14) imply that a classical particle incident in the direction of $\mathbf{p}_0$ with impact parameter b undergoes specular reflection at one of the two discontinuities and emerges in the direction of observation $\mathbf{p}$.

The important point about the conditions (2.14) is that, however large we make the cutoff radius ρ, there are always two values of the impact parameter b that satisfy the conditions. And the larger we make ρ, the larger the corresponding values of b. We can evaluate the contribution to the cross section (2.13) of this scattering, using the stationary-phase method to write down $\delta\psi_{\rho,b}$ in the relevant regions. The results of this rather tedious calculation are as follows: First, the second term in (2.13) makes no contribution to σ_ρ as $\rho \to \infty$. (This is because ψ^{Coul} is appreciable only when b is small, while $\delta\psi$ is appreciable only when b is close to $\rho\cos\theta/2$.) Second, the last term in (2.13) picks up an appreciable contribution to σ_ρ only from those b in the neighborhoods of the values in (2.14), and the contribution from each of these two neighborhoods is exactly

$\frac{1}{4}\sigma_{\text{Coul}}$ (in the limit that $\rho \to \infty$). Since the first term in (2.13) is already equal to σ_{Coul}, we conclude that the complete cross section for large screening radius is

$$\lim_{\rho \to \infty} \sigma_\rho(d\Omega - \phi) = \tfrac{3}{2}\sigma_{\text{Coul}}(d\Omega - \phi)$$

as already derived by a different method.

To summarize: We have proved the following results in Born approximation. The cross section σ_ρ for the sharply cutoff Coulomb potential does not approach the pure Coulomb cross section however large we make the cutoff radius ρ. The scattering of packets with small impact parameters b contributes exactly σ_{Coul}; but the specular reflection off the discontinuity of those packets with b near $\rho\cos\theta/2$ contributes a further unwanted $\frac{1}{2}\sigma_{\text{Coul}}$. This second term is not present for smoothly screened Coulomb potentials, and for such potentials σ_ρ does approach σ_{Coul} as $\rho \to \infty$.

3. SCREENED SHORT RANGE POTENTIALS

We next consider a short-range, central potential $V(r)$ satisfying

$$V(r) = O(r^{-3-\epsilon}) \quad \text{as } r \to \infty,$$

$$V(r) = O(r^{-2+\epsilon}) \quad \text{as } r \to \infty, \tag{3.1}$$

$V(r)$ piecewise continuous for $0 < r < \infty$.

We define the screened potential

$$V_\rho(r) = V(r)\,\alpha_\rho(r),$$

where we now require only that

$\alpha_\rho(0) = 1$,

$\alpha_\rho(r) \to 0$ monotonically as $r \to \infty$ (with ρ fixed)

$\alpha_\rho(r) \to 1$ as $\rho \to \infty$ (with r fixed).

$$\tag{3.2}$$

We shall prove that as $\rho \to \infty$ the exact cross section σ_ρ for V_ρ converges to the exact cross section σ for V,

$$\lim_{\rho \to \infty} \sigma_\rho(d\Omega - \phi) = \sigma(d\Omega - \phi) \tag{3.3}$$

We should perhaps emphasize that in this section we work with the exact cross sections, not just those of the Born approximation.

Some comments on our assumptions are in order. Our result can actually be proved for nonspherical potentials, but the proof is more complicated and needs some additional technical assumptions on the potentials.[19] For simplicity, therefore, we confine attention here to central potentials. Concerning the screening function $\alpha_\rho(r)$ it will be seen that we require somewhat less than the conditions (1.3) used in the Coulomb case. In the Coulomb case we had to require that $\alpha_\rho(r)$ be $O(r^{-2-\epsilon})$ as $r \to \infty$ in order that the screened potential be of short range; in the present case this is unnecessary since $V(r)$ is itself of short range. It will be seen that we do not require any kind of smoothness for the screening function. Therefore, our proof includes screening functions that are not smooth, like the sharp cutoff $\alpha_\rho(r) = \theta(\rho - r)$.

We begin by noting that Dollard's work[14] can be easily extended to all potentials and screening functions satisfying the conditions (3.1) and (3.2). Thus it follows

that the scattering operator S_ρ for V_ρ converges strongly to that for V as $\rho \to \infty$

$$\text{s-}\lim_{\rho \to \infty} S_\rho = S. \tag{3.4}$$

Now the scattering probability $W(d\Omega \leftarrow \phi)$ can be written as

$$W(d\Omega \leftarrow \phi) = \int_{d\Omega} d^3p \, |S\phi(\mathbf{p})|^2,$$

where the integral extends over the semi-infinite cone defined by $d\Omega$. From (3.4) it immediately follows that the probability W_ρ for scattering off the screened potential V_ρ converges to that for V as $\rho \to \infty$

$$\lim_{\rho \to \infty} W_\rho(d\Omega \leftarrow \phi) = W(d\Omega \leftarrow \phi). \tag{3.5}$$

As we have seen, the result (3.5) does not by itself guarantee the corresponding result for the cross sections. Rather than trying to prove the latter starting from (3.5), we shall prove it directly. We use the observation, made in connection with Eq. (2.3), that the screened cross section σ_ρ certainly converges to σ, if the screened amplitude $f_\rho(p\mathbf{u} \leftarrow \mathbf{p})$ converges to the unscreened amplitude $f(p\mathbf{u} \leftarrow \mathbf{p})$

$$\lim_{\rho \to \infty} f_\rho(p\mathbf{u} \leftarrow \mathbf{p}) = f(p\mathbf{u} \leftarrow \mathbf{p}) \tag{3.6}$$

uniformly for all $\mathbf{p}$ in any compact region not containing the origin $\mathbf{p} = 0$. We shall prove (3.6) by examining the partial-wave series for f_ρ and f.

The amplitude $f_\rho(p\mathbf{u} \leftarrow \mathbf{p})$ is given by a partial-wave series

$$f_\rho(p\mathbf{u} \leftarrow \mathbf{p}) = \sum (2l + 1) f_\rho^l(p) P_l(\cos\theta) \tag{3.7}$$

with a similar series for f in terms of partial-wave amplitudes f^l. Martin[20,21] has shown that the partial wave amplitudes satisfy bounds of the form

$$|f_\rho^l(p)| \leq \text{const } p^{-2} \int_0^\infty dr \, |j_l(pr)|^2 |V_\rho(r)| \tag{3.8}$$

for l sufficiently large. Since $|V_\rho(r)| \leq |V(r)|$ we can replace $|V_\rho(r)|$ by $|V(r)|$ in this bound and obtain a bound which is uniform in ρ. The integral in (3.8) can be evaluated to give

$$|f_\rho^l(p)| \leq \text{const } l^{-2-\epsilon} \tag{3.9}$$

for all ρ and l and for all p in any finite interval. The unscreened amplitudes $f^l(p)$ satisfy the same bound.

It follows from (3.9) that the partial-wave series (3.7) for $f_\rho(p\mathbf{u} \leftarrow \mathbf{p})$ is uniformly convergent for all ρ and all $\mathbf{p}$ in any compact region. Similarly the partial wave series for the unscreened amplitude $f(p\mathbf{u} \leftarrow \mathbf{p})$ is uniformly convergent for all $\mathbf{p}$ in the same region. Thus, to prove (3.6), it is sufficient to prove that, as $\rho \to \infty$ for any fixed l, the screened partial-wave amplitude $f_\rho^l(p)$ converges to the unscreened amplitude $f^l(p)$

$$\lim_{\rho \to \infty} f_\rho^l(p) = f^l(p) \tag{3.10}$$

uniformly for p in any finite interval not including $p = 0$.

We can prove (3.10) using Calogero's variable phase method[22] (much as in Ref. 8 for the Coulomb case). We let $\delta_\rho^l(r)$ denote the phase function of angular momentum l for the screened potential V_ρ. We first note that be-

cause $V(r)$ is short range [as in (3.1)], the screened phase shift can be written as

$$\delta_\rho^l = \delta_\rho^l(R) + O(R^{-2-\epsilon}), \tag{3.11}$$

where the estimate $O(R^{-2-\epsilon})$ is uniform for all ρ and all p in the relevant intervals. There is a similar expression for the unscreened δ^l. Secondly, one can check that for any fixed R and as $\rho \to \infty$, the screened phase function $\delta_\rho^l(R)$ approaches the unscreened $\delta^l(R)$

$$\lim_{\rho \to \infty} \delta_\rho^l(R) = \delta^l(R) \tag{3.12}$$

again uniformly for all p in the relevant intervals. Combining (3.11) and (3.12), we see that

$$\lim_{\rho \to \infty} \delta_\rho^l = \delta^l,$$

and hence that the corresponding limit holds for the partial-wave amplitudes f_ρ^l and f^l as required.

[1]W. Gordon, Z. Phys. **48**, 180 (1928).

[2]M. L. Goldberger and K. M. Watson, *Collision Theory* (Wiley, New York, 1964).

[3]J. R. Taylor, *Scattering Theory* (Wiley, New York, 1972).

[4]J. D. Dollard, J. Math. Phys. **5**, 729 (1964).

[5]E. Prugovecki and J. Zorbas, Nucl. Phys. A **213**, 541 (1973).

[6]E. Prugovecki and J. Zorbas, J. Math. Phys. **14**, 1398 (1973).

[7]E. Prugovecki and J. Zorbas, J. Phys. A **7**, 1557 (1974).

[8]J. R. Taylor, Nuovo Cimento B **23**, 313 (1974).

[9]M. D. Semon and J. R. Taylor, Nuovo Cimento A **26**, 48 (1975).

[10]In Refs. 8, 9 we assumed that the function $\int_0^{2\pi} d\varphi \, \phi(p)$ is twice differentiable in $\cos\theta$. This is certainly true if $\phi(p)$ is infinitely differentiable in rectangular coordinates (see Ref. 11).

[11]M. D. Semon, "Scattering by Long-Range and Screened Potentials," Ph. D. thesis, Univ. of Colorado (1976).

[12]That is, we assume that the cone defined by the incident packet does not overlap the observation cone defined by $d\Omega$.

[13]For notational convenience, we work with the cross section $\sigma(d\Omega \leftarrow \phi)$, which is the usual differential cross section times the solid angle, $\sigma(d\Omega \leftarrow \phi) = (d\sigma/d\Omega)d\Omega$.

[14]J. D. Dollard, J. Math. Phys. **9**, 620 (1968).

[15]Dollard gives his proof for exponential screening function, but it can be extended to all screening functions and potentials which satisfy the conditions given in Sec. 3.

[16]E. Campesino and J. R. Taylor, J. Math. Phys. **16**, 1227 (1975).

[17]In Ref. 16 it was assumed that the potential is $L^2(R^3)$, and is locally Hölder continuous except at a finite number of point singularities. It can be shown that the proof can be extended to include spherically symmetric potentials which satisfy conditions (3.1). As we mention below, (2.2) is also convergent when W is replaced by W_{Coul}, computed with the conventional Coulomb amplitude.

[18]Our precise conditions should probably not be taken too seriously since we are working in Born approximation. It seems quite possible that more stringent conditions would be needed in the exact case.

[19]In this case we require that the potentials satisfy the conditions (2.1) of Ref. 16. See Ref. 11.

[20]A. Martin, Nuovo Cimento **23**, 641 (1962).

[21]Martin's proof is reported in De Alfaro and Regge, *Potential Scattering* (North-Holland, Amsterdam, 1965), pp. 84—85.

[22]F. Calogero, *The Variable Phase Approach to Potential Scattering* (Academic, New York, 1967).

Branching rules and Clebsch–Gordan series for E_8 *

W. McKay and J. Patera

Centre de Recherches Mathématiques, Université de Montréal, Montréal, Canada

R. T. Sharp

Department of Physics, McGill University, Montreal, Canada
(Received 16 March 1976)

Branching rules for the second lowest dimensional representation of the exceptional simple Lie algebra of type E_8 are given with repsect to all its 14 maximal semisimple subalgebras. This representation of E_8 is of dimension 3875, the maximal subalgebras are of types A_8, D_8, $A_1 * A_7$, $A_1 * E_7$, $A_2 * E_6$, $A_3 * D_5$, $A_4 * A_4$, $A_1 * A_2 * A_5$, $G_2 * F_4$, $A_1 * A_2$, C_2, and three nonconjugate subalgebras all of type A_1. The Clebsch–Gordan series, necessary for decomposition of the direct product of three representations of dimension 248, are given.

I. INTRODUCTION

The purpose of this paper is to present the solution of two computational problems arising in some recent group-theoretical applications in particle physics. More precisely, we reduce the second lowest dimensional representation of the exceptional simple Lie algebra of type E_8 (this representation is of dimension 3875) to the representations of all its 14 maximal semisimple subalgebras by finding explicitly the irreducible representations of each of these subalgebras contained in the representation of E_8. Such a reduction is often called a "branching rule." We also find a decomposition of four direct products of representations of E_8 into direct sums of irreducible components (Clebsch–Gordan series), which are necessary for a complete decomposition of the direct product of three lowest dimensional representations (dimension 248) of E_8.

Our computation is motivated by a class of models currently being considered as possible classification schemes for elementary partices.[1] The relevant feature of these models is the use of the simple exceptional Lie groups generated by the Lie algebras of types G_2, F_4, E_6, E_7, and E_8. It is well known from past experience with particle symmetries, that the branching rules as well as reduction of direct products of representations are indispensable mathematical tools for such applications. Our results are also of interest in pure mathematics, for instance, as intermediate steps in some questions related to finite groups over fields of characteristic different from zero.[2]

The branching rules for the lower representations of all exceptional simple Lie algebras but E_8 are known. Indeed, in Ref. 3 all their representations of dimension less than 1000 are reduced to representations of all maximal semisimple subalgebras. Among the nontrivial representations of E_8 only the lowest one has dimension smaller than 1000 so that its reduction is contained in Ref. 3. Clearly this is insufficient in any application. The next lowest dimension of an irreducible representation of E_8 equals 3875. The branching rules for this representation are published here for the first time.

The method of our branching rule computation is that of Ref. 4 implemented as a computer program[5] and used in a simpler version in Ref. 3. Our results are summarized in Table II of Sec. 2. An independent verification of the table was done by checking the equality of dimensions, second-, and fourth-order indices of representations. General properties of these indices are described in a separate paper.[6]

Much simpler is the problem of the Clebsch—Gordan series. There, at least in principle, exists the Kostant—Steinberg formula[7] which solves the problem in general. For the cases of interest here it is, however, far simpler to guess the solution just from the equality of dimensions and indices. Table IV of Sec. 3 summarizes the four Clebsch—Gordan series needed to decompose the direct product of E_8 representations $(248) \otimes (248) \otimes (248)$.

In the Appendix a table of a few supplementary branching rules is presented, which together with Table II and the tables of Ref. 3 yield the branching rules of the E_8 representation of dimension 3875 with respect to any semisimple subalgebra of E_8, not necessarily maximal.

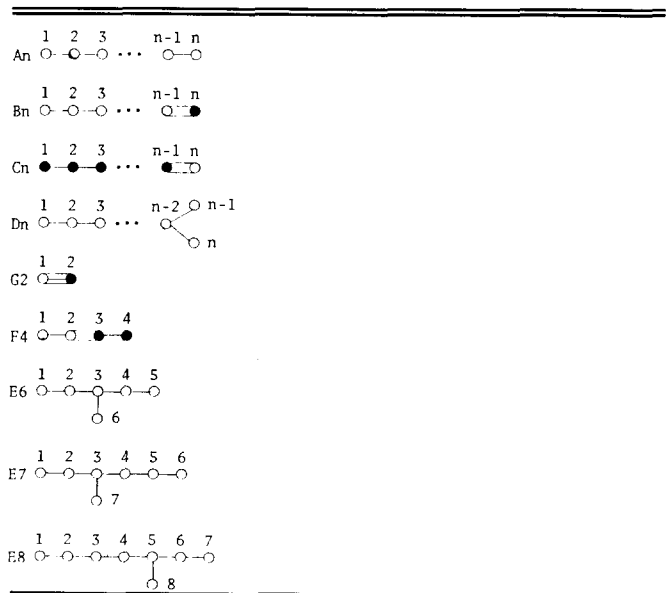TABLE I. Numbering of simple roots. Black dots represent shorter roots.

TABLE II. Branching rules for representations (10000000) and (00000010) of E_8 with respect to maximal semisimple subalgebras. Irreducible representations belonging to the same reducible representation of each subalgebra are arranged vertically.

```
E8                                    E8
(10000000)                            (00000010)

A8           G2*F4        A8           A1*A2*A5     A3*D5        A4*A4        C2
(00100000)   (01)(0001)   (10001000)   (1)(10)(01001) (010)(00100) (1010)(0100) 2(44)
(00000100)   (00)(1000)   (01000010)   (1)(01)(10010) (100)(10010) (1001)(1001) (81)
(10000001)   (10)(0000)   (00010001)   (0)(10)(10100) (001)(10001) (0101)(0010) (43)
                          (11000000)   (0)(01)(00101) (101)(01000) (0100)(0101) (80)
D8           A1*A2        (00000011)   (0)(00)(01010) (000)(00011) (0100)(0101) (24)
(00000001)   (2)(22)      (10000001)   (1)(11)(00100) (110)(00010) (1100)(0001) (61)
(01000000)   (4)(30)                   (0)(11)(10001) (011)(00001) (1000)(1100) (05)
             (4)(03)      D8           (0)(11)(10001) (000)(20000) (0011)(1000) 2(42)
A1*A7        (6)(11)      (00010000)   (1)(00)(11000) (200)(10000) (0001)(0011) (23)
(0)(0001000) (0)(11)      (10000010)   (2)(10)(00010) (002)(10000) (1100)(0001) (04)
(0)(1000001) (2)(00)      (20000000)   (2)(10)(00010) (100)(00001) (2000)(1000) 2(41)
(1)(0100000)              (2)(00)(10001) (010)(10000) (0110)(0000) (22)
(1)(0000010) C2           A1*A7        (0)(20)(01000) (101)(00000) (1000)(0002) (03)
(2)(0000000) (60)         (0)(0100010) (0)(10)(00002) (020)(00000) (0002)(0001) 2(40)
             (23)         (1)(1000100) (0)(02)(00010) (000)(00000) (0001)(2000) (02)
A2*E6        (20)         (1)(0010010) (0)(01)(20000)              (0000)(0110) (01)
(00)(000001)              (0)(1010000) (1)(20)(00001) A1           (1000)(0010)
(10)(100000) A1           (0)(0000101) (1)(02)(10000) (92)         (0100)(1000) A1
(01)(000010) (58)         (2)(0001000) (1)(00)(00100) (84)         (0010)(0001) (60)
(11)(000000) (46)         (2)(1000001) (0)(10)(00010) (80)         (0001)(0100) (56)
             (38)         (1)(2000000) (0)(01)(01000) (78)         (0000)(1001) (54)
A3*D5        (34)         (1)(0000002) (0)(00)(10001) (76)         (0000)(0000) 2(52)
(001)(00010) (26)         (0)(1000001) 2(1)(10)(10000) 2(72)                   (50)
(100)(00001) (22)         (1)(0100001) 2(1)(01)(00001) (70)        A1           3(48)
(010)(10000) ( 2)         (1)(0000010) (2)(11)(00000) 2(68)       (72)         2(46)
(000)(01000)              (0)(0000000) (0)(11)(00000) (66)        (68)         4(44)
(101)(00000) A1           A1*E7        2(0)(00)(00000) 2(64)      (64)         3(42)
             (46)         (0)(0000100)              2(64)         (62)         5(40)
A4*A4        (38)         (1)(0000001) A1*A2        (62)          2(60)        5(38)
(1000)(0010) (34)         (2)(1000000) 2( 8)(22)    3(60)         (58)         7(36)
(0100)(1000) (28)         (1)(0000000) ( 6)(41)     2(58)         2(56)        5(34)
(0010)(0001) (26)         (0)(0000000) ( 6)(14)     3(56)         2(54)        8(32)
(0001)(0100) (22)                      ( 4)(33)     2(54)         3(52)        6(30)
(1001)(0000) (18)         A1*E7        (10)(11)     3(52)         2(50)        9(28)
(0000)(1001) (14)         (0)(0000100) ( 6)(22)     2(50)         4(48)        7(26)
             (10)         (1)(0000001) ( 4)(41)     4(48)         3(46)        10(24)
A1*A2*A5                  (2)(1000000) ( 4)(14)     3(46)         5(44)        7(22)
(0)(00)(10001)           (0)(0000000) ( 2)(33)     4(44)         3(42)        10(20)
(0)(01)(01000) A1                      ( 8)(11)     2(42)         5(40)        7(18)
(1)(00)(00100) (38)      A2*E6        2( 6)(30)    4(40)         4(38)        9(16)
(0)(10)(00010) (34)      (01)(010000) 2( 6)(03)    3(38)         7(36)        6(14)
(1)(10)(10000) (28)      (10)(000100) 3( 4)(22)    5(36)         4(34)        9(12)
(1)(01)(00001) (26)      (00)(100010) ( 2)(41)     3(34)         6(32)        4(10)
(0)(11)(00000) (22)      (11)(000001) ( 2)(14)     4(32)         5(30)        6( 8)
(2)(00)(00000) 2(22)     (20)(000010) ( 0)(60)     2(30)         7(28)        3( 6)
             (18)        (02)(100000) ( 0)(06)     4(28)         5(26)        5( 4)
A1*E7                    (10)(100000) 2( 6)(11)    3(26)         7(24)        3( 0)
(0)(1000000) (16)        (01)(000010) ( 4)(30)     5(24)         4(22)
(1)(0000010) (14)        (11)(000000) ( 4)(03)     2(22)         7(20)
(2)(0000000) (10)        (00)(000000) 2( 2)(22)    4(20)         4(18)
             ( 6)                     3( 4)(11)    (18)          6(16)
             ( 2)        G2*F4        2( 2)(30)    3(16)         3(14)
                         (01)(0010)   2( 2)(03)    2(14)         6(12)
                         (10)(1000)   2( 0)(22)    4(12)         2(10)
                         (00)(0002)   ( 6)(00)     2( 8)         4( 8)
                         (02)(0001)   2( 2)(11)    2( 4)         ( 6)
                         (01)(0001)   ( 4)(00)     2( 0)         3( 4)
                         (02)(0000)   ( 0)(11)                   2( 0)
                         (00)(0000)   ( 0)(00)
```

II. BRANCHING RULES

We describe the notations in Tables I and II. They coincide with those of Ref. 3. Since our results are straightforward reproductions of computer outputs, they do not contain any lower or upper indices. Thus simple Lie algebras are denoted as $E8$, $E7$, $G2$, $A8$, etc. An irreducible representation of a simple Lie algebra of rank n is specified by n components a_i as $(a_1 a_2 \cdots a_n)$, where

$$a_i = 2(\Lambda, \alpha_i)/(\alpha_i, \alpha_i), \quad i = 1, 2, \ldots, n. \tag{1}$$

In (1), α_i denotes the ith simple root of the corresponding algebra. For each algebra the simple roots are numbered as in Table I. The brackets (,) denote a scalar product; Λ is the highest weight of the representation. The components a_i are known to be nonnegative integers.

A representation of a semisimple algebra which is a product of several simple ones is written as $(a_1 a_2 \cdots a_{n_1})$ $(b_1 b_2 \cdots b_{n_2}) \cdots$, where each bracket characterizes an irreducible representation of a simple algebra in the product and $n_1, n_2, \cdots$ are corresponding ranks. When-

TABLE III. Dimensions[8], second-, fourth-order indices for some representations of E_8.

Representation	Dimension	$I^{(2)}$	$I^{(4)}$
(10000000)	248	480	960
(01000000)	30380	117600	517440
(00100000)	2450240	14227200	96744960
(00010000)	146325270	1132840800	10422135360
(00001000)	6899079264	66765283200	774477285120
(00000100)	6696000	42336000	314979840
(00000010)	3875	12000	41280
(00000001)	147250	684000	3666240
(30000000)	1763125	10920000	79497600
(20000000)	27000	108000	492480
(11000000)	4096000	24576000	173015040
(00000020)	4881384	31492800	239345280
(10000010)	779247	4021920	24131520
(10100000)	344452500	2755620000	26233502400

TABLE IV. Clebsch—Gordan Series for E_8.

Direct products	Direct sums
(10000000) × (10000000)	(20000000) + (01000000) + (00000010) + (10000000) + (0000000)
(10000000) × (20000000)	(30000000) + (11000000) + (10000010) + (01000000) + (20000000) + (10000000)
(10000000) × (01000000)	(11000000) + (00100000) + (10000010) + (00000001) + (0100000) + (20000000)
	+ (00000010) + (10000000)
(10000000) × (00000010)	(10000010) + (00000001) + (01000000) + (00000010) + (10000000)

ever the multiplicity of an irreducible representation in a direct sum exceeds one, it is written in front of the representation.

We remark that representations of algebras A_1 and A_2 sometimes have the component a_i equal to a two digit number. When this happens for A_1 algebras, no misunderstanding is possible, in the case of A_2 (cf. Table VI) a comma separates a_1 from a_2.

III. CLEBSCH-GORDAN SERIES

Dimensions and indices of relevant representations of E_8 are summarized in Table III.[8]

Our problem is to find the decomposition of the direct product of three lowest representations of E_8,

$$(10000000) \times (10000000) \times (10000000), \qquad (2)$$

into a direct sum. First we decompose the direct product of two, (10000000)×(10000000), into a direct sum and then multiply each of its components by (10000000). Corresponding Clebsch—Gordan series are found in Table IV.

The results of Table IV were obtained in a straightforward way by requiring equality of dimensions, second-, and fourth-indices. Thus for instance, (10000000)×(01000000) has dimension $N_1 N_2 = 248 \cdot 30\,380 = 7\,534\,240$, the second index[6] $N_1 I_2^{(2)} + N_2 I_1^{(2)} = 248 \cdot 117\,600 + 30\,380 \cdot 480 = 43\,747\,200$, and the fourth index[6] $N_1 I_2^{(4)} + N_2 I_1^{(4)} + \frac{5}{2} I_1^{(2)} I_2^{(2)} = 248 \cdot 517\,440 + 30\,380 \cdot 960 + \frac{5}{2} 480 \cdot 117\,600 = 298\,609\,920$. Here $N, I^{(2)}, I^{(4)}$ denote dimension, second index, and fourth index respectively; their values are taken from Table III. These numbers must respectively be equal to the sums of dimensions, second indices and fourth indices of the irreducible components contained in the corresponding direct sum. Just looking at Table III one readily concludes that our result in Table IV is the only way to satisfy all three requirements.

In every direct product of two irreducible representations one can easily find the two highest irreducible components of the direct sum. Indeed, it is well known that the first component is obtained just by adding the representation labels. Thus in the example above we get the representation (11000000) of E_8. The highest weight of the second component is equal to[9]

$$\Lambda_1 + \Lambda_2 - (\alpha_{i_1} + \alpha_{i_2} + \cdots), \qquad (3)$$

where Λ_1 and Λ_2 are the highest weights of the repre-

sentations which are being multiplied and the expression in the brackets is the minimal chain[9] of simple roots connecting Λ_1 and Λ_2. In our example above we have

$$\Lambda_1 = (10000000)$$

$$= 2\alpha_1 + 3\alpha_2 + 4\alpha_3 + 5\alpha_4 + 6\alpha_5 + 4\alpha_6 + 2\alpha_7 + 3\alpha_8,$$

$$\Lambda_2 = (01000000)$$

$$= 3\alpha_1 + 6\alpha_2 + 8\alpha_3 + 10\alpha_4 + 12\alpha_5 + 8\alpha_6 + 4\alpha_7 + 6\alpha_8, \qquad (4)$$

$$\alpha_{i_1} + \alpha_{i_2} + \cdots = \alpha_1 + \alpha_2.$$

Hence the highest weight of the second component in the direct sum is

$$4\alpha_1 + 8\alpha_2 + 12\alpha_3 + 15\alpha_4 + 18\alpha_5 + 12\alpha_6 + 6\alpha_7 + 9\alpha_8$$

$$= (00100000). \qquad (5)$$

APPENDIX

The purpose of this appendix is to supplement Table II in such a way that branching rules for the E_8 representation of dimension 3875 with respect to any semisimple subalgebra of E_8 (not necessarily a maximal one) can be read off directly from tables. For that we have to find the branching rules for all irreducible representations of subalgebras of E_8 which occur in Table II and whose dimensions are equal to or exceed 1000. Indeed, for the remaining ones the branching rules are found in Ref. 3. Consequently, we have to find the branching rules for the representations of Table V, with the exception of (00010001) of A_8 which is contragredient to (10001000).

TABLE V. Dimensions $(N \geq 1000)$, second-, and fourth-order indices of irreducible representations of semisimple subalgebras of E_8 contained in (00000010) of E_8.

Algebra	Representation	Dimension	2-index	4-index
$A8$	(10001000)	1050	3360	11 760
$A8$	(01000010)	1215	3888	13 824
$A8$	(00010001)	1050	3360	11 760
$D8$	(00010000)	1820	5824	20 608
$D8$	(10000010)	1920	5888	19 968
$E7$	(0000100)	1539	4536	15 120
$B7$	(0001000)	1365	4004	13 244
$D7$	(0001000)	1001	3080	10 640

TABLE VI. Branching rules for representations of Table V with respect to all their maximal semisimple subalgebras.

A8 (10001000)	A8 (01000010)	D8 (10000010)	D9 (00010000)	E7 (0000100)		B7 (0001000)	D7 (0001000)

A8 (10001000)

```
A7
(1001000)
(1000100)
(0001000)
(0000100)

B4
(1002)
(0002)

A2*A2
(11)(22)
(30)(30)
(22)(11)
(11)(03)
(11)(30)
(30)(11)
(03)(11)
(00)(22)
2(11)(11)
(22)(00)
(00)(11)
(11)(00)

A1*A6
(1)(100100)
(0)(101000)
(0)(100010)
(2)(000100)
(1)(001000)
(0)(000100)
(1)(000010)

A2*A5
(01)(10100)
(10)(10010)
(11)(00100)
(00)(11000)
(20)(00010)
(01)(00010)
(10)(01000)
(00)(10001)
(00)(00100)
(10)(00001)

A3*A4
(010)(1010)
(110)(0010)
(001)(1100)
(101)(0100)
(100)(1001)
(200)(0001)
(001)(0010)
(010)(0001)
(000)(2000)
(100)(1000)
(000)(0100)
(000)(1000)
(100)(0000)
```

A8 (01000010)

```
A7
(0100010)
(1000010)
(0100001)
(1000001)

B4
(0200)
(1010)
(0002)

A2*A2
(11)(22)
(30)(03)
(03)(30)
(22)(11)
(11)(03)
(11)(30)
(03)(11)
(30)(11)
(00)(22)
3(11)(11)
(22)(00)
(00)(11)
(11)(00)
(00)(00)

A1*A6
(0)(010010)
(1)(100010)
(1)(010001)
(2)(100001)
(0)(000010)
(0)(010000)
(1)(000001)
(0)(000000)

A2*A5
(00)(01010)
(10)(10010)
(01)(01001)
(11)(10001)
(01)(00010)
(10)(01010)
(00)(10001)
(02)(00001)
(20)(10000)
(01)(10000)
(10)(00001)
(00)(00000)

A3*A4
(101)(1001)
(001)(0101)
(100)(1010)
(000)(0110)
(011)(0001)
(110)(1000)
(010)(0010)
(010)(0100)
(020)(0000)
(000)(1001)
(100)(0001)
(001)(1000)
(101)(0000)
(000)(0000)
```

D8 (10000010)

```
A7
(1000100)
(0000101)
(1010000)
(0010001)
(0001000)
2(1000001)
(0000002)
(2000000)
(0000010)
(0100000)

B7
(1000001)
(0000001)

B4
(1002)
(1010)
(1100)
(0002)
(0100)
(1000)

A1*C4
(1)(1001)
(3)(1100)
(1)(1100)
(3)(0010)
(1)(0010)
(5)(1000)
(3)(1000)
(1)(1000)

A1*A1*D6
(0)(1)(100010)
(1)(0)(100001)
(1)(2)(000010)
(2)(1)(000001)
(0)(1)(000001)
(1)(0)(000010)

A3*D5
(001)(10010)
(100)(10001)
(011)(00010)
(110)(00001)
(001)(00001)
(100)(00010)

D4*D4
(1001)(0010)
(0001)(1010)
(1010)(0001)
(0010)(1001)
(0001)(0001)
(0010)(0010)

A1*B6
(1)(100001)
(3)(000001)
(1)(000001)

C2*B5
(10)(10001)
(11)(00001)
(10)(00001)

B3*B4
(001)(1001)
(101)(0001)
(001)(0001)

C2*C2
(30)(11)
(11)(30)
2(11)(11)
(12)(10)
(10)(12)
(30)(10)
(10)(30)
2(11)(10)
2(10)(11)
2(10)(10)
```

D9 (00010000)

```
A7
(0100010)
(0010001)
(1000100)
2(0001000)
(1000001)
(0100000)
(0000010)
(0000000)

B7
(0001000)
(0010000)

B4
(1002)
(0200)
(1100)
(0002)
(2000)

A1*C4
(2)(1010)
(0)(0200)
(4)(0001)
(4)(0100)
(2)(2000)
(2)(0100)
(4)(0000)
(0)(0000)

A1*A1*D6
(0)(0)(000100)
(1)(1)(001000)
(0)(2)(010000)
(2)(0)(010000)
(1)(1)(100000)
(0)(0)(000000)

A3*D5
(010)(00100)
(101)(01000)
(000)(00011)
(002)(10000)
(002)(10000)
(101)(00000)

D4*D4
(0100)(0100)
(0011)(1000)
(1000)(0011)
(0002)(0000)
(0020)(0000)
(0000)(0002)
(0000)(0020)

A1*B6
(0)(000100)
(2)(001000)
(2)(010000)
(0)(100000)

C2*B5
(00)(00010)
(01)(00100)
(20)(01000)
(20)(10000)
(01)(00000)

B3*B4
(100)(0010)
(010)(0100)
(000)(0002)
(002)(1000)
(002)(0000)

C2*C2
(21)(20)
(02)(02)
(20)(21)
(21)(01)
(01)(21)
(02)(01)
(40)(00)
2(20)(20)
(01)(02)
(00)(40)
(20)(01)
(01)(20)
(02)(00)
(01)(01)
(00)(02)
(01)(00)
(00)(01)
(00)(00)
```

E7 (0000100)

```
A7                    A1*G2
(0100010)             (4)(11)
(0010001)             (2)(03)
(0000101)             (0)(20)
(1000001)             (2)(11)
                      (6)(10)
E6                   2(4)(02)
(100010)              (2)(02)
(010000)              (6)(01)
(000100)             2(2)(10)
(000001)             2(0)(02)
2(100000)             (4)(01)
2(000000)             (2)(01)
                      (4)(00)
A2                    (0)(00)
( 6, 6)
(10, 1)               A1*A1
( 1,10)               (12)( 4)
( 5, 5)               (10)( 6)
( 3, 6)               ( 8)( 8)
( 6, 3)               (10)( 4)
( 5, 2)               ( 8)( 6)
( 2, 5)               ( 6)( 8)
( 3, 3)               (12)( 0)
( 2, 2)              2(10)( 2)
( 1, 1)              3( 8)( 4)
                     3( 6)( 6)
A1                   2( 4)( 8)
(52)                  ( 2)(10)
(48)                 2( 8)( 2)
2(44)                3( 6)( 4)
(42)                 2( 4)( 6)
2(40)                2( 8)( 0)
(38)                 4( 6)( 2)
3(36)                5( 4)( 4)
2(34)                3( 2)( 6)
4(32)                 ( 0)( 8)
2(30)                2( 4)( 2)
4(28)                3( 4)( 0)
3(26)                4( 2)( 2)
5(24)                2( 0)( 4)
5(20)                2( 0)( 0)
3(18)
5(16)                 A1*D6
2(14)                 (0)(000100)
5(12)                 (1)(100001)
2(10)                 (2)(010000)
4( 8)                 (0)(200000)
3( 4)                 (1)(000010)
2( 0)                 (0)(000000)

A1                    A2*A5
(40)                  (10)(10100)
2(36)                 (00)(01010)
(34)                  (01)(00101)
3(32)                 (11)(10001)
2(30)                 (01)(20000)
4(28)                 (20)(01000)
4(26)                 (02)(00010)
6(24)                 (10)(00002)
4(22)                 (00)(10001)
8(20)                 (11)(00000)
5(18)                 (00)(00000)
9(16)
5(14)                 A1*A3*A3
8(12)                 (0)(101)(101)
5(10)                 (1)(110)(100)
7( 8)                 (1)(100)(110)
3( 6)                 (1)(011)(001)
5( 4)                 (1)(001)(011)
3( 0)                 (0)(200)(010)
                      (2)(010)(010)
A1*F4                 (0)(002)(010)
(2)(0010)             (0)(010)(200)
(0)(0002)             (2)(000)(101)
(2)(1000)             (0)(010)(010)
(2)(0001)             (0)(000)(020)
(0)(0001)            2(1)(100)(001)
(4)(0000)            2(1)(001)(100)
(0)(0000)             (0)(101)(000)
                      (0)(000)(101)
G2*C3                2(0)(000)(000)
(02)(010)
(10)(200)
(01)(101)
(01)(200)
(00)(020)
(01)(010)
(02)(000)
(00)(002)
(01)(000)
(00)(010)
(00)(000)
```

B7 (0001000)

```
D7
(0001000)
(0010000)

A3
(113)
(040)
(311)
2(121)
(004)
2(202)
(400)
(012)
(210)
2(020)
2(101)

A1
(44)
(40)
(38)
2(36)
(34)
3(32)
2(30)
4(28)
3(26)
5(24)
4(22)
6(20)
4(18)
6(16)
4(14)
6(12)
3(10)
5( 8)
2( 6)
4( 4)
2( 0)

A1*C2
(2)(22)
(4)(40)
(6)(21)
(4)(21)
(0)(40)
(2)(21)
(4)(02)
(6)(20)
(8)(01)
(2)(02)
(4)(20)
(0)(02)
2(2)(20)
(4)(01)
(0)(01)
(4)(00)
(0)(00)

A1*D6
(0)(000100)
(2)(001000)
(2)(010000)
(0)(100000)

A1*A1*B5
(0)(0)(00010)
(1)(1)(00100)
(0)(2)(01000)
(2)(0)(01000)
(1)(1)(10000)
(0)(0)(00000)

C2*D5
(01)(00100)
(00)(00011)
(20)(01000)
(20)(10000)
(01)(00000)

A3*B4
(010)(0010)
(101)(0100)
(000)(0002)
(002)(1000)
(200)(1000)
(101)(0000)

B3*D4
(010)(0100)
(100)(0011)
(002)(1000)
(000)(0002)
(000)(0020)
(002)(0000)
```

D7 (0001000)

```
A6
(010010)
(001001)
(100100)
(000100)
(100001)
(001000)
(010000)
(000010)
(000000)

B6
(000100)
(001000)

C3
(111)
(400)
(210)
(020)
(101)
(010)

C2
(61)
(05)
(42)
(23)
(41)
(22)
(03)
(40)
(21)
(02)
(01)

G2
(13)
(04)
(12)
(20)
(11)
(02)
(10)

A1*A1*D5
(1)(1)(00100)
(0)(0)(00011)
(0)(2)(01000)
(2)(0)(01000)
(1)(1)(10000)
(0)(0)(00000)

A3*D4
(101)(0100)
(010)(0011)
(002)(1000)
(200)(1000)
(000)(0002)
(000)(0020)
(101)(0000)

A1*B5
(0)(00010)
(2)(00100)
(2)(01000)
(0)(10000)

C2*B4
(01)(0010)
(00)(0002)
(20)(0100)
(20)(1000)
(01)(0000)

B3*B3
(010)(010)
(002)(100)
(100)(002)
(002)(000)
(000)(002)
```

The branching rules for all representations of Table V with respect to all maximal subalgebras were obtained using the same computer program. The results are summarized in Table VI.

[1]F. Gürsey, P. Ramond, and P. Sikivie, Phys. Rev. D **12**, 2166 (1975); F. Gürsey, *Proc. Kyoto Int. Symp. on Math. Problems in Theor. Phys.*, edited by H. Araki (Springer, New York, 1975); F.Gürsey, Proc. 4th Int. Colloq. on Group Theor. Methods in Physics, Nijmegen, 1975, Holland (to be published); F. Gürsey, P. Ramond, and P. Sikivie, Phys. Lett. B **60**, 177 (1976); L.C. Biedenharn, Proc. of the Symposium on Mathematical Physics, Mexico City, 1976 (to be published); F. Gürsey and P. Sikivie, Phys. Rev. Lett. **36**, 775 (1976); F. Gürsey, Proc. of Orbis Scientiae, Coral Gables, 1976 (to be published).

[2]R.W. Carter, *Simple groups of Lie type* (Wiley-Interscience, New York, 1972).

[3]J. Patera and D. Sankoff, *Tables of branching rules for representations of simple Lie algebras* (Presses Université de Montréal, Montreal, 1973).

[4]A. Navon and J. Patera, J. Math. Phys. **8**, 489 (1967).

[5]W. McKay, J. Patera, and D. Sankoff, "The computation of branching rules for representations of semisimple Lie algebras," in *Computers in the Study of Nonassociative Rings and Algebras*, edited by R.E. Beck and B. Kolman (Academic, New York, 1976) (to be published).

[6]J. Patera, R.T. Sharp, and P. Winternitz, preprint CRM-602 (Centre Rech. Math., Université de Montréal, 1976).

[7]B. Kostant, Trans. Am. Math. Soc. **93**, 53 (1959).

[8]Dimension of the following E_8 representations are shown incorrectly in Table 30 of Ref. 9: (01000000), (00100000), (00010000), and (00001000). A similar error is in Table 5 of Ref. 10 where the 2-indices of these representations are shown (our 2-index should differ from the index of Dynkin by a factor equal to the rank of the algebra).

[9]E.B. Dynkin, Tr. Mosk. Math. Obsc. **1**, 39 (1952); Am. Math. Soc. Transl. Ser. 2 **6**, 245 (1965).

[10]E.B. Dynkin, Mat. Sb. N.S. **30**, 349 (1952); Am. Math. Soc. Transl. Ser. 2 **6**, 111 (1965).

1375   J. Math. Phys., Vol. 17, No. 8, August 1976

McKay, Patera, and Sharp   1375

The state labeling problems for SO(*N*) in U(*N*) and U(*M*) in Sp(2*M*)

H. S. Green, C. A. Hurst, and Y. Ilamed*

Department of Mathematical Physics, University of Adelaide, S. Australia 5001
(Received 20 August 1975)

It is shown that, in a boson representation, the operators whose eigenvalues serve to label representations of SO(*N*) in U(*N*) also serve to label representations of U(*M*) in Sp(2*M*). The problem of labeling U(2) in Sp(4) is considered in detail, and it is shown how to find labeling operators with rational eigenvalues, depending, however, on the representation. The solution of this problem is shown to provide a solution of the equivalent problem of the labeling of SO(3) in U(3).

1. INTRODUCTION

The study of the representations of the classical groups has, in recent years, been motivated as much by important physical applications as by its intrinsic mathematical interest. Most of the associated problems have been solved, in principle at least, but several problems remain which, in spite of their apparent mathematical simplicity and physical importance, have not yielded to persistent attack. One of these, formulated by Racah[1] and Ilamed,[2] concerns the definition of an operator with known eigenvalues, to complete the labeling of irreducible representations of SO(3) within U(3). Hughes[3] and Judd, Miller, Patera, and Winternitz[4] have recently shown how to determine the eigenvalues of two different operators, which however turn out to be irrational in general, and no general formula for the eigenvalues is known. Green and Bracken,[5] on the other hand, have introduced an operator with integral eigenvalues, probably related to the integral parameter of Bargmann and Moshinsky[6]; but so far no explicit definition of this operator has been found. There is a similar problem, of some importance in relativistic quantum mechanics, concerning the labeling of irreducible representations of SO(4) within U(4); this has been considered in a similar way by Jarvis,[7] but again only a partial solution has been obtained. This is, of course, true also of the more general problem of labeling representations of SO(*N*) within U(*N*).

An apparently unrelated problem concerns the labeling of the states of Sp(2*M*) with operators related to the integral parameters of the Weyl or Gel'fand bases. Using boson representations of the generators, Lohe and Hurst[8] have considered the problem of labeling Sp(2*M* − 2) in Sp(2*M*), but again no explicit labeling operators have been found in general.

Finally, it may be mentioned that Govorkov[9] has encountered an apparently intractable problem (for *p* ≥ 3) when seeking labeling operators for representations of U(*N*) within generalized parafermion algebras of order *p*. There is, besides, an analogous problem, not yet discussed in the literature, associated with the labeling of representations of generalized paraboson algebras.

Our intention is to show that all the problems mentioned above are closely related, and, as usual, pro-gress made in the solution of one of the problems is an important aid to the solution of the others. We shall make use of boson realizations of the generators of the algebras (or, equivalently, of differential operators), and the reciprocal relation between representations of Sp(2*M*) and U(*N*) which has been exploited recently by Quesne and Moshinsky.[10] It is easy, in this formalism, to show that the problem of labeling irreducible representations of SO(*N*) within U(*N*) is equivalent, for a suitable choice of *M*, to the problem of labeling irreducible representations of U(*M*) within Sp(2*M*). We establish simple relations between the invariants of the algebras concerned, and show that, for *N* = 3 or 4 and *M* = 2, the solution of the problem may be found within an interesting algebra, which is not a finite dimensional Lie algebra, though it has finite-dimensional representations. It is shown how to compute matrix elements of all the invariants of SO(3) or SO(4) within U(3) or U(4), equivalently of U(2) within Sp(4), and hence to determine their eigenvalues.

2. TENSOR REPRESENTATIONS OF U(*N*) AND Sp(2*M*)

In canonical form, the generators b_{ij} of U(*N*) satisfy the commutation relations

$$[b_{ij}, b_{kl}] = \delta_{kj} b_{il} - \delta_{il} b_{kj}, \quad (i,j,k,l = 1, \ldots, N). \tag{1}$$

Irreducible representations may be labeled by eigenvalues of the first *N* of the invariants

$$\langle b \rangle = b_{ii}, \quad \langle b^2 \rangle = b_{ij} b_{ji},$$
$$\langle b^3 \rangle = b_{ij} b_{jk} b_{ki}, \tag{2}$$

etc., or of the set of invariants $(L_1, L_2, \ldots, L_N)$ whose eigenvalues in finite dimensional representations are the highest weights. They are related to the $\langle b^r \rangle$ by

$$\sum_{r=1}^{N} L_r = \langle b \rangle,$$
$$\sum_{r=1}^{N} L_r (L_r + N + 1 - 2r) = \langle b^2 \rangle, \tag{3}$$

and similar but more complicated identities of higher degree. The (eigenvalues of the) L_r differ by integers, and $L_1 \geq L_2 \geq \cdots \geq L_N$. If

$$C_{ij} = b_{ij} + c \delta_{ij}, \tag{4}$$

where *c* is a constant, the C_{ij} are also generators of

$U(N)$, with representations labeled $(L_1 + c,\ L_2 + c,\ \cdots,\ L_N + c)$.

The generators l_{ij} of the orthogonal subgroup SO(N) of U(N) may be defined by

$$l_{ij} = b_{ij} - b_{ji}, \tag{5}$$

or $l = b - \bar{b}$, where $\bar{b}$ denotes the transpose of the matrix b whose elements are b_{ij}. If $H = [\tfrac{1}{2}N]$ (i.e., $\tfrac{1}{2}N$ when N is even, but $\tfrac{1}{2}N - \tfrac{1}{2}$ when N is odd), irreducible representations of SO(N) may be labeled by eigenvalues of the first H of the invariants

$$\langle l^2 \rangle = 2\langle b^2 - b\bar{b}\rangle = l_{ij}l_{ji},$$
$$\langle l^4 \rangle = \langle (\bar{b} - b)^4 \rangle = l_{ij}l_{jk}l_{kp}l_{pi}, \tag{6}$$

etc., or of the set of invariants $(l_1, l_2, \ldots, l_H)$ corresponding to the highest weights. They are related to the $\langle l^{2s} \rangle$ by

$$2 \sum_{s=1}^{H} l_s(l_s + N - 2s) = \langle l^2 \rangle \tag{7}$$

and similar identities of higher degree.[11]

Next we consider Sp($2M$). If we denote the generators by S_{PQ} $(P, Q = 1, 2, \ldots, 2M)$ the commutation relations

$$[S_{PQ}, S_{UV}] = g_{UP}S_{QV} + g_{UQ}S_{PV} + g_{VP}S_{QU} + g_{VQ}S_{PU} \tag{8}$$

are satisfied, where $S_{PQ} = S_{QP}$ but $g_{PQ} = -g_{QP}$. If $S^P{}_Q = g^{PR}S_{RQ}$, where $g^{PR}g_{RQ} = \delta^P{}_Q$, the invariants $\langle S^2 \rangle$, ..., $\langle S^{2M} \rangle$ defined by

$$\langle S^2 \rangle = S^P{}_Q S^Q{}_P,$$
$$\langle S^4 \rangle = S^P{}_Q S^Q{}_U S^U{}_V S^V{}_P, \tag{9}$$

etc., may serve to label irreducible representations. Alternatively, the set of invariants $(\Lambda_1, \Lambda_2, \ldots, \Lambda_M)$, whose eigenvalues in finite dimensional irreducible representations are the highest weights, may be used: these are related to the $\langle S^{2p} \rangle$ by[11]

$$2 \sum_{S=1}^{M} \Lambda_S(\Lambda_S + 2M + 2 - 2S) = \langle S^2 \rangle \tag{10}$$

and similar identities of higher degree. We may choose

$$g_{PQ} = g^{QP} = \delta_{P\ Q+M} - \delta_{P+M\ Q}. \tag{11}$$

If λ and μ take integral values between 1 and M, we define

$$\alpha_{\lambda\mu} = S_{\lambda\mu}, \quad \alpha^{\lambda\mu} = S_{\lambda+M\ \mu+M},$$
$$\alpha^{\lambda}{}_{\mu} = S_{\lambda+M\ \mu}, \tag{12}$$

so that the commutation relations (8) reduce to

$$[\alpha^{\lambda}{}_{\mu}, \alpha^{\nu}{}_{\rho}] = \delta^{\nu}{}_{\mu}\alpha^{\lambda}{}_{\rho} - \delta^{\lambda}{}_{\rho}\alpha^{\nu}{}_{\mu},$$
$$[\alpha^{\lambda}{}_{\mu}, \alpha^{\nu\rho}] = \delta^{\nu}{}_{\mu}\alpha^{\lambda\rho} + \delta^{\rho}{}_{\mu}\alpha^{\lambda\nu},$$
$$[\alpha_{\lambda\mu}, \alpha^{\nu}{}_{\rho}] = \delta^{\nu}{}_{\lambda}\alpha_{\mu\rho} + \delta^{\nu}{}_{\mu}\alpha_{\lambda\rho},$$
$$[\alpha_{\lambda\mu}, \alpha^{\nu\rho}] = \delta^{\nu}{}_{\lambda}\alpha^{\rho}{}_{\mu} + \delta^{\rho}{}_{\lambda}\alpha^{\nu}{}_{\mu} + \delta^{\nu}{}_{\mu}\alpha^{\rho}{}_{\lambda} + \delta^{\rho}{}_{\lambda}\alpha^{\rho}{}_{\mu}. \tag{13}$$

The elements $\alpha^{\lambda}{}_{\mu}$ are evidently generators of the unitary subgroup U(M) of Sp($2M$), and the invariants of this subgroup analogous to the $\langle b^r \rangle$ are $\langle \alpha \rangle$, ..., $\langle \alpha^M \rangle$, where

$$\langle \alpha \rangle = \alpha^{\lambda}{}_{\lambda}, \quad \langle \alpha^2 \rangle = \alpha^{\lambda}{}_{\mu}\alpha^{\mu}{}_{\lambda},$$
$$\langle \alpha^3 \rangle = \alpha^{\lambda}{}_{\mu}\alpha^{\mu}{}_{\nu}\alpha^{\nu}{}_{\lambda}, \tag{14}$$

etc. The set of invariants $(\lambda_1, \lambda_2, \ldots, \lambda_M)$, analogous to $(l_1, l_2, \ldots, l_N)$, are given by

$$\sum_{r=1}^{M} \lambda_r = \langle \alpha \rangle,$$
$$\sum_{r=1}^{M} \lambda_r(\lambda_r + M + 1 - 2r) = \langle \alpha^2 \rangle \tag{15}$$

and similar identities of higher degree.[11]

We now introduce a set of boson creation and annihilation operators [or coordinates and differential operators], denoted by a_{Pi} $(P = 1, \ldots, 2M;\ i = 1, \ldots, N)$, and satisfying

$$[a_{Pi}, a_{Qj}] = g_{QP}\delta_{ij}. \tag{16}$$

We may suppose that

$$a_{\lambda+Mi} = a^*_{\lambda i} \quad [\text{or } \partial/\partial a_{\lambda i}], \quad \lambda \leqslant M' \leqslant M,$$
$$a_{\lambda i} = -a^*_{\lambda+Mi} \quad [\text{or } -\partial/\partial a_{\lambda+Mi}], \quad M' < \lambda, \tag{17}$$

where a^*_{Pi} now means the Hermitian conjugate of a_{Pi}. Then, if

$$a^{\lambda}{}_i = a_{\lambda+Mi},$$

$a^{\lambda}{}_i$ will be a creation operator [or coordinate] for $\lambda \leqslant M'$, and an annihilation or differential operator for $\lambda > M'$.

Products of creation operators like $a^{\lambda}{}_i a^{\mu}{}_j \cdots a_{\nu k}$ $a_{\rho l} \cdots$ (where $\lambda, \mu, \cdots \leqslant M'$ and $\nu, \rho, \cdots > M'$) can be regarded as vectors of (reducible) representations of either U(N) or Sp($2M$). In such representations, the generators of U(N) are

$$b_{ij} = a^{\lambda}{}_i a_{\lambda j} + \tfrac{1}{2}M\delta_{ij}, \tag{18}$$

with an appropriate choice of the constant c in (4), and those of Sp($2M$) are

$$S_{PQ} = a_{Pi}a_{Qi} + \tfrac{1}{2}Ng_{PQ}, \tag{19}$$
$$\alpha^{\lambda}{}_{\mu} = a^{\lambda}{}_i a_{\mu i} + \tfrac{1}{2}N\delta^{\lambda}{}_{\mu}, \tag{20}$$
$$\alpha^{\lambda\mu} = a^{\lambda}{}_i a^{\mu}{}_i, \quad \alpha_{\lambda\mu} = a_{\lambda i}a_{\mu i}.$$

The commutation relations (1) and (13) are satisfied on account of (16), or

$$[a_{\lambda i}, a^{\mu}{}_j] = \delta^{\mu}{}_{\lambda}\delta_{ij}. \tag{21}$$

Since each generator b_{ij} of U(N) commutes with each generator $\alpha^{\lambda}{}_{\mu}$ of the unitary subgroup U(M) of Sp($2M$), the invariants of U(M) are also invariants of U(N) in this representation. It is, indeed, easy to see that

$$\langle \alpha \rangle = a^{\lambda}{}_i a_{\lambda i} + \tfrac{1}{2}MN = \langle b \rangle,$$
$$\langle (\alpha - \tfrac{1}{2}N)(\alpha + \tfrac{1}{2}N) \rangle = a^{\lambda}{}_i a_{\mu i}a_{\lambda j}a^{\mu}{}_j = \langle (b - \tfrac{1}{2}M)(b + \tfrac{1}{2}M) \rangle,$$

and, more generally, it has been shown by M.C.K. Aguilera-Navarro and V.C. Aguilera-Navarro

$$\langle (\alpha - \tfrac{1}{2}N)(\alpha + \tfrac{1}{2}N)^n \rangle = \langle (b - \tfrac{1}{2}M)(b + \tfrac{1}{2}M)^n \rangle,$$
$$\langle (\bar{\alpha} + \tfrac{1}{2}N)(\bar{\alpha} - \tfrac{1}{2}N)^n \rangle = \langle (\bar{b} + \tfrac{1}{2}M)(\bar{b} - \tfrac{1}{2}M)^n \rangle, \tag{22}$$

for $n = 0, 1, 2, \cdots$, where $\bar{\alpha}$ again means the transpose of the matrix α with elements $\alpha^{\lambda}{}_{\mu}$ [so that, e.g., $(\bar{\alpha}^2)^{\lambda}{}_{\mu} = \bar{\alpha}^{\lambda}{}_{\nu}\bar{\alpha}^{\nu}{}_{\mu} = \alpha^{\nu}{}_{\lambda}\alpha^{\mu}{}_{\nu}$]. A short proof of the results of (22), and others needed below, is given for convenience in the Appendix to this paper. It follows from (22), together

1377   J. Math. Phys., Vol. 17, No. 8, August 1976

Green, Hurst, and Ilamed   1377

with (15) and (3), that the invariants $(\lambda_1, \lambda_2, \ldots, \lambda_M)$ of U(M) are connected with those, $(L_1, L_2, \ldots, L_N)$, of U(N) by

$$\lambda_r' - \tfrac{1}{2}N = L_r' - \tfrac{1}{2}M, \quad r \leqslant \min(M, N),$$

$$\lambda_r' - \tfrac{1}{2}N = 0, \quad N < r \leqslant M, \tag{23}$$

$$L_r' - \tfrac{1}{2}M = 0, \quad M < r \leqslant N,$$

where λ_r' and L_r' are defined in terms of the λ_r and L_r by relations of the type $\lambda_r' - r = \lambda_s - s$, $L_r' - r = L_s - s$, chosen so that the eigenvalues of both $(\lambda_1, \lambda_2, \ldots, \lambda_M)$ and $(L_1, L_2, \ldots, L_N)$ are in decreasing order.

Similarly, since each generator S_{PQ} of Sp($2M$) commutes with each generator l_{ij} of the orthogonal subgroup SO(N) of U(N) in the representation considered, the invariants of SO(N) are also invariants of Sp($2M$). It may be verified explicitly that

$$\langle (S - \tfrac{1}{2}N)(S + \tfrac{1}{2}N) \rangle = a^P{}_i a_{Qi} a_{Pj} a^Q{}_j = \langle (l - M)(l + M + 1) \rangle$$

and, as shown in the Appendix,

$$\langle (S - \tfrac{1}{2}N)(S + \tfrac{1}{2}N)^n \rangle = \langle (l - M)(l + M + 1)^n \rangle \tag{24}$$

for $n = 0, 1, 2, \cdots$. It follows from this, together with (10) and (17), that the invariants $(\Lambda_1, \Lambda_2, \ldots, \Lambda_M)$ of Sp($2M$) and those, $(l_1, l_2, \ldots, l_H)$, of O(N) are connected by

$$\Lambda_s' + M + 1 - \tfrac{1}{2}N = l_s, \quad s \leqslant \min(H, M),$$

$$\Lambda_s' + M + 1 - \tfrac{1}{2}N = 0, \quad H < s \leqslant M,$$

$$l_s' = 0, \quad M < s \leqslant H, \tag{25}$$

where Λ_s' and l_s' are defined in terms of the Λ_s and l_s by relations of the type $\Lambda_s' - s = \Lambda_r - r$, $l_s' - s = l_r - r$, chosen so that the eigenvalues of both $(\Lambda_1, \Lambda_2, \ldots, \Lambda_M)$ and $(l_1, l_2, \ldots, l_H)$ are in decreasing order.

The operators whose eigenvalues could serve to label equivalent irreducible representations of SO(N) within an irreducible representation of U(N) are also related to the operators whose eigenvalues could serve to label equivalent representations of U(M) within an irreducible representation of Sp($2M$). To construct such operators, we introduce linear operators $A, \bar{A}, B,$ and $\bar{B}$, defined on arbitrary symmetric tensors $\phi_{\lambda\mu}$, $\phi^{\lambda\mu}$, and ψ_{ij} by

$$\phi_{\lambda\mu}A = \phi_{\lambda\nu}\alpha^\nu{}_\mu + \phi_{\mu\nu}\alpha^\nu{}_\lambda,$$

$$\phi^{\lambda\mu}\bar{A} = \phi^{\lambda\nu}\alpha^\mu{}_\nu + \phi^{\mu\nu}\alpha^\lambda{}_\nu,$$

$$B\psi_{ij} = b_{ik}\psi_{kj} + b_{jk}\psi_{ki}, \tag{26}$$

$$\bar{B}\psi_{ij} = b_{ki}\psi_{kj} + b_{kj}\psi_{ki}.$$

The tensor $(\bar{b}b + \bar{b})_{ij} = b_{ki}b_{kj} + b_{ji}$ is symmetric, and it is readily verified that

$$\langle \bar{\alpha}_c \alpha_c \rangle = \alpha^{\lambda\mu}\alpha_{\lambda\mu} = \langle (\bar{b} - \tfrac{1}{2}M - 1)(b - \tfrac{1}{2}M) \rangle,$$

$$\langle \bar{\alpha}_c \alpha_c (A + N) \rangle = \alpha^{\lambda\mu}(2\alpha_{\lambda\nu}\alpha^\nu{}_\mu + N\alpha_{\lambda\mu})$$

$$= \langle (\bar{b} - \tfrac{1}{2}M - 1)(b - \tfrac{1}{2}M)(2b + M) \rangle$$

$$= \langle (\bar{b} - \tfrac{1}{2}M - 1)(b - \tfrac{1}{2}M)(B + M) \rangle,$$

and, more generally, as shown in the Appendix,

$$\langle \bar{\alpha}_c \alpha_c (A + N)^n \rangle = \langle (\bar{b} - \tfrac{1}{2}M - 1)(b - \tfrac{1}{2}M)(B + M)^n \rangle. \tag{27}$$

Similarly,

$$\langle \alpha_c \bar{\alpha}_c \rangle = \alpha_{\lambda\mu}\alpha^{\lambda\mu} = \langle (b + \tfrac{1}{2}M + 1)(\bar{b} + \tfrac{1}{2}M) \rangle$$

and, as shown in the Appendix,

$$\langle \alpha_c \bar{\alpha}_c (\bar{A} - N)^n \rangle = \langle (\bar{b} + \tfrac{1}{2}M + 1)(\bar{b} + \tfrac{1}{2}M)(\bar{B} - M)^n \rangle. \tag{28}$$

These, and similar results which can be derived by the same method for expressions involving more than one pair of factors $\alpha_{\lambda\mu}, \alpha^{\nu\rho}$, show that a solution to the problem of labeling representations of SO(N) within an irreducible representation of SU(N) is also a solution to the problem labeling representations of U(M) within a corresponding irreducible representation of Sp($2M$), and conversely. However, as can be seen from (23) and (25), the most general representation of SU(N) can only be realized by taking $M \geqslant N - 1$, while the most general representation of Sp($2M$) can only be realized by taking $H = [\tfrac{1}{2}N] \geqslant M$. In particular, the solution of the problem of labeling general representations of SU(3) is solved by taking $M = 2$, but the solution of the problem of labeling general representations of Sp(4) is solved by taking $N = 4$.

The representations of Sp($2M$) in terms of boson operators, constructed in this section, are necessarily infinite dimensional; finite dimensional representations could be obtained, if desired, by using fermion or parafermion operators[13] instead. The infinite dimensionality may be thought of as associated with the multiplicity of *distinct* representations of U(M) contained within an irreducible representation of Sp($2M$). The operators whose eigenvalues serve to label *equivalent* representations of U(M) within Sp($2M$) are, as we have seen, the same as those which label equivalent representations of SO(N) within U(N), and generate an algebra $\underline{A}$ with finite dimensional representations. For $M = 2$, we shall investigate the structure of this algebra below. For the sake of symmetry, we eventually choose $M' = 1$ in (17), so that U(2) is strictly replaced by the noncompact form U(1,1). However, analogous results hold also for $M' = 0$ and $M' = 2$, so that the conclusions do not depend in any essential way on this choice.

3. LABELING OF U(2) IN Sp(4)

An irreducible representation of Sp(4) is labeled by the eigenvalues of the invariants (Λ_1, Λ_2), related to those defined in (9) by

$$\langle S^2 \rangle = 2(\Lambda^2 + \Lambda'^2 - 5),$$

$$\langle S^4 \rangle = 2(\Lambda^4 + \Lambda'^4 + 3\Lambda^2 + 3\Lambda'^2 - 32), \tag{29}$$

$$\Lambda = \Lambda_1 + 2 = l_1 + \tfrac{1}{2}N - 1,$$

$$\Lambda' = \Lambda_2 + 1 = l_2 + \tfrac{1}{2}N - 2,$$

where l_1 and l_2 take nonnegative integral eigenvalues. Distinct representations of U(2) in an irreducible representation of Sp(4) are labeled by eigenvalues of (λ_1, λ_2), related to the invariants defined in (14) by

$$\langle \alpha \rangle = \lambda - \lambda', \quad \langle \alpha^2 \rangle = \lambda^2 + \lambda'^2 - \tfrac{1}{2},$$

$$\lambda = \lambda_1 + \tfrac{1}{2}, \quad \lambda' = -\lambda_2 + \tfrac{1}{2}. \tag{30}$$

If $M' = 1$ in (17), the following inequalities are satisfied:

$$\lambda \geq \tfrac{1}{2}(N-1), \quad \lambda' \geq \tfrac{1}{2}(N-1),$$

$$\lambda + \lambda' \geq l' = l_1 + l_2 + N - 1, \tag{31}$$

so that if $N=3$ and $l_2=0$, λ and λ' are positive integers such that $\lambda + \lambda' \geq l_1 + 2$, but if $N=4$ they are half-odd-integers $\geq \tfrac{3}{2}$ such that $\lambda + \lambda' \geq l_1 + l_2 + 3$.

Equivalent representations of U(2) within an irreducible representation of Sp(4) are distinguished by eigenvalues of elements of the algebra $\underline{A}$ of invariants of U(2), constructed from the tensors $\alpha^{\lambda\mu}$ and $\alpha_{\lambda\mu}$, as well as the generators of U(2). In order to determine the representations of the algebra $\underline{A}$, it is helpful to introduce also a set of operators $(P, \bar{P}, Q, \bar{Q}, R, \bar{R}, S, \bar{S})$ which shift from one irreducible representation of U(2) to another. The simplest of these operators can be defined directly by

$$R = \alpha^\lambda{}_1 \alpha_{\lambda 2} - \alpha^\lambda{}_2 \alpha_{\lambda 1}, \quad \bar{R} = \alpha^{\lambda 2}\alpha^1{}_\lambda - \alpha^{\lambda 1}\alpha^2{}_\lambda,$$

$$S = \alpha_{11}\alpha_{22} - (\alpha_{12})^2 = \epsilon^{\lambda\mu}\epsilon^{\rho\sigma}\alpha_{\lambda\rho}\alpha_{\mu\sigma},$$

$$S = \alpha^{11}\alpha^{22} - (\alpha^{12})^2 = \epsilon_{\lambda\mu}\epsilon_{\rho\sigma}\alpha^{\lambda\rho}\alpha'^{\mu\sigma}. \tag{32}$$

They can be regarded as nonvanishing elements of antisymmetric tensors, and therefore change the eigenvalues of (λ, λ'), as defined in (30), by $(-1, +1)$, $(+1, -1)$, $(-2, +2)$, and $(+2, -2)$, respectively. The remaining shift operators can be constructed from symmetric tensors, and are easily derived by making use of the characteristic identities [11]

$$(A - \lambda + \lambda' - 2)(A - 2\lambda - 1)(A + 2\lambda' - 1) = 0,$$
$$(\bar{A} - \lambda + \lambda' + 2)(\bar{A} - 2\lambda + 1)(\bar{A} + 2\lambda' + 1) = 0, \tag{33}$$

satisfied by the linear operators A and $\bar{A}$ introduced in (26), when $M=2$. By the omission of one of the factors of the left sides of these identities, we obtain projection operators which, applied to $\alpha_{\lambda\mu}$ and $\alpha^{\lambda\mu}$, isolate the required shift operators. Thus, if

$$\rho_j = \lambda + \lambda' + j \tag{34}$$

and

$$\alpha_{\lambda\mu} = \alpha^+_{\lambda\mu}(\rho_{-1}\rho_0)^{-1} - 2\alpha^0_{\lambda\mu}(\rho_{-1}\rho_1)^{-1} + \alpha^-_{\lambda\mu}(\rho_0\rho_1)^{-1},$$

$$\alpha^+_{\lambda\mu} = \tfrac{1}{2}\alpha_{\lambda\mu}(A - \lambda + \lambda' - 2)(A + 2\lambda' - 1),$$

$$\alpha^0_{\lambda\mu} = \tfrac{1}{2}\alpha_{\lambda\mu}(A - 2\lambda - 1)(A + 2\lambda' - 1), \tag{35}$$

$$\alpha^-_{\lambda\mu} = \tfrac{1}{2}\alpha_{\lambda\mu}(A - \lambda + \lambda' - 2)(A - 2\lambda - 1),$$

then any component of the symmetric tensors $\alpha^+_{\lambda\mu}$, $\alpha^0_{\lambda\mu}$, and $\alpha^-_{\lambda\mu}$ will change (λ, λ') by $(-2, 0)$, $(-1, +1)$, and $(0, +2)$, respectively. Similarly, if

$$\alpha^{\lambda\mu} = \alpha_+^{\lambda\mu}(\rho_{-1}\rho_0)^{-1} - 2\alpha_0^{\lambda\mu}(\rho_{-1}\rho_1)^{-1} + \alpha_-^{\lambda\mu}(\rho_0\rho_1)^{-1},$$

$$\alpha_+^{\lambda\mu} = \tfrac{1}{2}\alpha^{\lambda\mu}(\bar{A} - \lambda + \lambda' + 2)(\bar{A} - 2\lambda + 1),$$

$$\alpha_0^{\lambda\mu} = \tfrac{1}{2}\alpha^{\lambda\mu}(\bar{A} + 2\lambda' + 1)(\bar{A} + 2\lambda + 1), \tag{36}$$

$$\alpha_-^{\lambda\mu} = \tfrac{1}{2}\alpha^{\lambda\mu}(\bar{A} - \lambda + \lambda' + 2)(\bar{A} + 2\lambda + 1),$$

then any component of $\alpha_+^{\lambda\mu}$, $\alpha_0^{\lambda\mu}$, and $\alpha_-^{\lambda\mu}$ will change (λ, λ') by $(0, -2)$, $(+1, -1)$, and $(+2, 0)$, respectively. Clearly $\alpha^0_{\lambda\mu}$ and $\alpha_0^{\lambda\mu}$ must differ from R and $\bar{R}$ in (32) only by factors depending on the U(2) generators; and all the required shift operators, except S and $\bar{S}$ can be defined by

$$\alpha_{11}^+ = P(\gamma_0\gamma_1\rho_0/\rho_2)^{1/2}, \quad \alpha_-^{11} = \bar{P}(\gamma_{-2}\gamma_{-1}\rho_0/\rho_{-2})^{1/2},$$

$$\alpha_{11}^0 = -R\alpha^2{}_1, \quad \alpha_0^{11} = -\bar{R}\alpha^1{}_2,$$

$$\alpha_{11}^- = Q(\alpha^2{}_1)^2(\rho_0/\gamma_0\gamma_1\rho_{-2})^{1/2}, \quad \alpha_+^{11} = \bar{Q}(\alpha^1{}_2)^2(\rho_0/\gamma_{-2}\gamma_{-1}\rho_2)^{1/2}$$

$$\gamma_j = \lambda - \alpha^1{}_1 + j + \tfrac{1}{2} = \lambda' + \alpha^2{}_2 + j + \tfrac{1}{2}, \tag{37}$$

with normalization factors chosen so that

$$P\bar{P} = \alpha_{\lambda\mu}\alpha_+^{\lambda\mu}, \quad \bar{P}P = \alpha^{\lambda\mu}\alpha_{\lambda\mu}^-,$$

$$R\bar{R} = \alpha_{\lambda\mu}\alpha_0^{\lambda\mu}, \quad \bar{R}R = \alpha^{\lambda\mu}\alpha_{\lambda\mu}^0, \tag{38}$$

$$Q\bar{Q} = \alpha_{\lambda\mu}\alpha_-^{\lambda\mu}, \quad \bar{Q}Q = \alpha^{\lambda\mu}\alpha_{\lambda\mu}^+.$$

Commutation relations for the shift operators so defined can be obtained from the commutation relations satisfied by $\alpha_{\lambda\mu}$ and $\alpha^{\lambda\mu}$, by substituting from (33) and (34) into (13) and separating terms which shift between one irreducible representation of U(2) and another. Thus we obtain

$$[P, Q] = 4\rho_0 S, \quad [P, \bar{Q}] = 0,$$

$$[P, R] = 0, \quad [P, \bar{R}] = [\bar{Q}, R],$$

$$[P, S] = [Q, S] = [R, S] = 0,$$

$$[S, \bar{R}] = 2(\lambda - \lambda' + 1)R,$$

$$[S, \bar{P}] = 4\lambda Q, \quad [S, \bar{Q}] = 4\lambda' P,$$

$$QP = R^2 + \rho_1^2 S, \tag{39}$$

together with conjugate relations, like $[\bar{Q}, \bar{P}] = 4\rho_0 \bar{S}$.

All elements of the algebra A can be expressed as functions of $\lambda, \lambda', \Lambda, \Lambda'$, and two independent invariants

$$X = \tfrac{1}{2}[R, \bar{R}], \quad Y = \tfrac{1}{2}(R\bar{R} + \bar{R}R). \tag{40}$$

Obviously $R\bar{R} = X + Y$ and $\bar{R}R = Y - X$, but if we make use of the explicit expressions for $\langle S^2 \rangle$ and $\langle S^4 \rangle$ in (9), and the above relations (39), we also obtain

$$\bar{P}P = Y + \rho_{-2}X + \rho_{-1}^2\phi_{-1}, \quad P\bar{P} = Y + \rho_2 X + \rho_1^2\phi_1,$$

$$\bar{Q}Q = Y - \rho_2 X + \rho_1^2\phi_1', \quad Q\bar{Q} = Y - \rho_{-2}X + \rho_{-1}^2\phi_{-1}',$$

$$\bar{S}S = Y - (\sigma - 2)X + \phi_{-1}\phi_1' + K, \quad S\bar{S} = Y - (\sigma + 2)X + \phi_1\phi_{-1}' + K,$$

$$\bar{S}R^2 + \bar{R}^2 S = (\phi_{-2} + \phi_2' - \rho_0^2 + 1)Y - X^2 + (\sigma - 4)X + (1 - \rho_0^2)K,$$

$$S\bar{R}^2 + R^2\bar{S} = (\phi_2 + \phi_{-2}' - \rho_0^2 + 1)Y - X^2 + (\sigma + 4)X + (1 - \rho_0^2)K, \tag{41}$$

where

$$\phi_j = (\lambda + j)^2 - \tfrac{1}{2}(\Lambda^2 + \Lambda'^2 - \tfrac{1}{2}),$$

$$\phi_j' = (\lambda' + j)^2 - \tfrac{1}{2}(\Lambda^2 + \Lambda'^2 - \tfrac{1}{2}), \quad \sigma = \lambda - \lambda',$$

$$K = \tfrac{1}{2}(\Lambda^2 + \Lambda'^2 - \tfrac{1}{2}) - \tfrac{1}{4}(\Lambda^2 - \Lambda'^2)^2. \tag{42}$$

It also follows from the commutation relations that

$$[X, Y] = 2(\bar{R}^2 S - \bar{S}R^2) = 2(S\bar{R}^2 - R^2\bar{S}). \tag{43}$$

Clearly the operators X and Y do not commute in general, and are not elements of any finitely generated Lie algebra. However, there are several special classes of representations of U(2) in which X and Y commute, and in fact have unique eigenvalues. Since for $M' = 1$, $\lambda + \lambda' \geq l'$, according to (31), if $|r\rangle$ is a vector belonging to a representation of U(2) such that $\lambda + \lambda' = l'$ or $l' + 1$, we must have $P|r\rangle = \bar{Q}|r\rangle = 0$; it then follows from (41) that X and Y have eigenvalues

1379   J. Math. Phys., Vol. 17, No. 8, August 1976

Green, Hurst, and llamed   1379

$$X = -\tfrac{1}{2}\rho_{-1}{}^2\sigma, \quad Y = -\tfrac{1}{2}\rho_{-1}{}^2(\dot\phi_{-1} + \phi'_{-1})$$

$$(\lambda + \lambda' = l' \text{ or } l' + 1), \tag{44}$$

in such representations. Also, when λ has its minimum value $\tfrac{1}{2}(N-1)$, so that $\bar{P}P = \bar{S}S = 0$, and when λ' has its minimum value $\tfrac{1}{2}(N-1)$, or next to minimum value $\tfrac{1}{2}(N+1)$, so that $\bar{Q}Q = S\bar{S} = 0$, X and Y have unique eigenvalues given by

$$X = Y = -\rho_{-1}\phi_{-1}, \qquad [\lambda = \tfrac{1}{2}(N-1)],$$

$$X - 2\phi_{-2} = Y + 2\rho_{-2}\phi_{-2} = -\rho_{-1}\phi_{-1}, \quad [\lambda = \tfrac{1}{2}(N+1)],$$

$$X = -Y = \rho_{-1}\phi'_{-1}, \qquad [\lambda' = \tfrac{1}{2}(N-1)],$$

$$X + 2\phi'_{-2} = Y + 2\rho_{-2}\phi'_{-2} = -\rho_{-1}\phi'_{-1} \quad [\lambda' = \tfrac{1}{2}(N+1)]. \tag{45}$$

To construct a general matrix representation for the operators $\bar{S}S$ and $S\bar{S}$, and hence for X and Y, we note that $\mu_1 = \lambda - \tfrac{1}{2}N + \tfrac{1}{2}$ and $\mu_2 = \lambda' - \tfrac{1}{2}N + \tfrac{1}{2}$ have nonnegative integral eigenvalues in irreducible representations of U(2) within the irreducible representation (Λ, Λ') of Sp(4), and that $\mu_1 + \mu_2$ is odd or even according as $\Lambda + \Lambda'$ is odd or even. Let us introduce a set of eigenvectors $|r\rangle_{p,q}$ of μ_1 and μ_2 such that

$$\mu_1|r\rangle_{p,q} = (r+2p)|r\rangle_{p,q}, \quad \mu_2|r\rangle_{p,q} = (s+2q)|r\rangle_{p,q}, \tag{46}$$

$$r + s = \Lambda + \Lambda', \quad |r\rangle_{p,q} = \bar{P}^p\bar{Q}^q|r\rangle_{0,0},$$

$$|r+2\rangle_{0,0} = [(r-\Lambda+1)^2 - 1/4]\bar{S}|r\rangle_{0,0}.$$

Clearly, in a representation of U(2) in which μ_1 and μ_2 have fixed eigenvalues, states are sufficiently labeled by r, which takes even or odd values ranging from $\max(\eta, \Lambda + \Lambda' - \mu_2)$ to $\min(\mu_1, \Lambda + \Lambda' - \eta)$ where $\eta = 0$ or 1 according as μ_1 is even or odd. From the commutation relations (39), and the known eigenvalues of X and Y in the states $|r\rangle_{0,0}$, given by (44), we find that

$$\bar{S}S|r\rangle_{p,q} = \{g(r-1)g(s+1) + (\mu_1{}^2 - r^2)[(\mu_2+2)^2 - s^2]\}|r\rangle_{p,q}$$

$$+ (\mu_1{}^2 - r^2)g(s-1)|r+2\rangle$$

$$+ [(\mu_2+2)^2 - (s+2)^2]g(r-1)|r-2\rangle,$$

$$S\bar{S}|r\rangle_{p,q} = \{g(r+1)g(s-1) + [(\mu_1+2)^2 - r^2](\mu_2{}^2 - s^2)\}$$

$$\times |r\rangle_{p,q} + (\mu_2{}^2 - s^2)g(r-1)|r-2\rangle$$

$$+ [(\mu_1+2)^2 - (r+2)^2]g(s-1)|r+2\rangle, \tag{47}$$

where

$$g(x) = (x - \Lambda') - \tfrac{1}{4}. \tag{48}$$

These formulas provide an explicit matrix representation for $\bar{S}S$ and $S\bar{S}$, and hence for X and Y, in an irreducible representation of U(2), labeled by $\lambda = \mu_1 + \tfrac{1}{2}N - \tfrac{1}{2}$ and $\lambda' = \mu_2 + \tfrac{1}{2}N - \tfrac{1}{2}$, within an irreducible representation of Sp(4), labeled by Λ and Λ'. From the matrices, eigenvalues can of course be computed without difficulty; they are irrational in general. In the next section we shall discuss the definition of an operator with eigenvalues corresponding to the integral parameter r.

4. LABELING OF SO(3) IN U(3)

The problem of finding an operator with known eigenvalues to label representations of SO(3) within an arbitrary irreducible representation of U(3) is the simplest and best known of the problems under consideration.

According to (23), if a representation of U(2) is labeled $(\lambda_1, \lambda_2) = (\mu_1 + N/2 - 1, -\mu_2 - N/2 + 1)$, the corresponding representation of U(N) will be labeled $(L_1, L_2, \ldots, L_N)$, where $L_1 = \mu_1$, $L_N = -\mu_2$, and the other L_r vanish. We therefore consider a representation of U(3) labelled $(\mu_1, 0, -\mu_2)$, with invariants

$$\langle b \rangle = \mu_1 - \mu_2 = \sigma,$$

$$\langle b^2 \rangle = \mu_1(\mu_1 + 2) + \mu_2(\mu_2 + 2). \tag{49}$$

The corresponding representation of SU(3) is labeled (μ_1, μ_2) and the invariants of SU(3) defined by Racah[1] and Ilamed[2] are

$$g = 2[\mu_1(\mu_1 + 2) + \mu_2(\mu_2 + 2) - (\mu_1 - \mu_2)^2/3],$$

$$g_3 = 4(\mu_1 - \mu_2)[5(\mu_1 - \mu_2)^2/9 - \mu_1(\mu_1 + 2)$$

$$- \mu_2(\mu_2 + 2) - 2]/3. \tag{50}$$

Different representations of SO(3) contained within an irreducible representation of U(3) of the type considered are labeled by the operator l, defined by

$$l(l+1) = \langle b^2 - b\bar{b} \rangle; \tag{51}$$

this is, of course, the angular momentum in quantum-mechanical applications. In the corresponding representations of Sp(4), the invariants defined in (9) are given by

$$\langle S^2 \rangle = 2l(l+1) - 9,$$

$$\langle S^4 \rangle = 2[(l+\tfrac{1}{2})^4 + 3(l+\tfrac{1}{2})^2 - 32 + \tfrac{13}{16}], \tag{52}$$

and the representations are labeled by

$$(\Lambda_1, \Lambda_2) = (l - \tfrac{3}{2}, -1 \pm \tfrac{1}{2}).$$

Operators whose eigenvalues, if they could be found, would serve to label equivalent representations of SO(3) in an irreducible representation of U(3), have been defined by Racah[1] and Ilamed[2]; given by

$$x = \langle b\bar{b}b + \bar{b}b\bar{b} - 2b \rangle + \langle b \rangle^3 + \langle b \rangle(4l(l+1)/3 - 3\langle b^2 \rangle),$$

$$y = 8\langle b \rangle x/3 - 4\langle b^2\bar{b}^2 \rangle - (16\langle b^2 \rangle/9 + 9)l(l+1)$$

$$+ 2\langle b^2 \rangle(\langle b^2 \rangle + 2\langle b \rangle^2 + 4) + \langle b \rangle^2 - 4 \tag{53}$$

in the present notation. The operators $O_l{}^0$ and $Q_l{}^0$ introduced by Hughes[3] are related to them by

$$O_l{}^0 = -3\sqrt{6}x,$$

$$Q_l{}^0 = -18y + 12l(l+1)[2\langle b^2 \rangle - 2\langle b \rangle^2/3 - l(l+1) - 3],$$

while those considered by Green and Bracken[5] are

$$S_3 = \langle b\bar{b}b + \bar{b}b\bar{b} \rangle,$$

$$S_4 = \langle b\bar{b}^2 b + \bar{b}b^2\bar{b} \rangle$$

$$= 2\langle b^2\bar{b}^2 \rangle + 9\langle b^2 \rangle - 3\langle b \rangle^2 - 12l(l+1).$$

Although the eigenvalues of these operators have been obtained in special representations,[3] they are irrational numbers in general, and no general formula is known. The algebra of the operators x and y was studied by Ilamed,[2] who has shown how to derive commutation relations of the type

$$[x, [x, y]] = 24y^2 + \zeta_1 y + \zeta_1 y + \zeta_2 x^2 + \zeta_3 x + \zeta_4,$$

$$[y, [y, x]] = 32x^3 + \zeta_2(xy + yx) + \zeta_3 y + \zeta_5 x^2 + \zeta_6 x + \zeta_7,$$

1380   J. Math. Phys., Vol. 17, No. 8, August 1976

Green, Hurst, and Ilamed   1380

$$[x, y]^2 = 16x^4 + \zeta_8 x^3 + 16y^3 + \zeta_9(x^2 y + yx^2) + \zeta_{10}(xy + yx)$$
$$+ \zeta_{11} x^2 + \zeta_{12} y^2 + \zeta_{13} x + \zeta_{14} y + \zeta_{15}, \tag{54}$$

where the ζ_i are given as simple polynomials in g, g_3, and

$$s = l(l+1). \tag{55}$$

Our object in this paper is to show how the results obtained in the previous section can be used to solve the above equations. As all the operators to be used are SO(3) invariants, we may consider an irreducible representation in which l has a fixed integral value. However, we shall find it convenient to consider a variety of representations of U(3), in which μ_1 and μ_2 take different nonnegative integral eigenvalues. The dimensions of the irreducible representations of x and y depend on the eigenvalues of μ_1 and μ_2 and are, as one can see from (46), never greater than $\frac{1}{2}(\mu_1 + \mu_2 - l) + 1$.

To solve (54), we first consider the representation for which $[x, y] = 0$. These can be found by setting the right sides of (54) equal to zero; as we have already seen in the last section, the solutions fall into six classes, all included in (44) or (45): (i) $\mu_1 + \mu_2 = l$, (ii) $\mu_1 + \mu_2 = l + 1$, (iii) $\mu_1 = 0$, (iv) $\mu_1 = 1$, (v) $\mu_2 = 0$, and (vi) $\mu_2 = 1$. The eigenvalues of x and y in these representations are related to those of X and Y found in the previous section, the precise relationship between the operators being given by

$$x = 2X + \sigma s/3,$$
$$y = -4Y + 4\sigma X/3 + (2\langle b^2 \rangle + 3 - 8\sigma^2/9)s. \tag{56}$$

Now, a general representation of U(3) is related to a corresponding representation of one of the classes (i) and (ii) listed above by a shift operator of the type $\bar{P}^{a-r} Q^{a+r}$, with $\bar{P}$ and Q defined as in the previous section. Thus, to determine a general representation of x and y, labeled by s, μ_1, μ_2, and r, it is sufficient to define $\bar{P}$ and Q. The considerations of the previous section show that the following factorizations are possible:

$$Y + (\mu_1 + \mu_2)X + (\mu_1 + \mu_2 + 1)^2 \phi_{-1} = \bar{P}P,$$
$$Y + (\mu_1 + \mu_2 + 4)X + (\mu_1 + \mu_2 + 3)^2 \phi_1 = P\bar{P},$$
$$Y - (\mu_1 + \mu_2 + 4)X + (\mu_1 + \mu_2 + 3)^2 \phi_1' = \bar{Q}Q,$$
$$Y - (\mu_1 + \mu_2)X + (\mu_1 + \mu_2 + 1)^2 \phi_{-1}' = Q\bar{Q},$$
$$\phi_j = (\mu_1 + j + 1)^2 - \tfrac{1}{2}s, \quad \phi_j' = (\mu_2 + j + 1)^2 - \tfrac{1}{2}s, \tag{57}$$

where P and $\bar{Q}$ are represented by rectangular matrices with two more rows than columns, and $\bar{P}$ and Q are their adjoints. By different factorizations,

$$Y + X = R\bar{R}, \quad Y - X = \bar{R}R,$$
$$Y - (\sigma - 2)X + \phi_{-1}\phi_1' + \tfrac{1}{2}s(1 - \tfrac{1}{2}s) = \bar{S}S,$$
$$Y - (\sigma + 2)X + \phi_1 \phi_{-1}' + \tfrac{1}{2}s(1 - \tfrac{1}{2}s) = S\bar{S}, \tag{58}$$

we can define the shift operators R and S and their adjoints. As shown by (47), a representation can be found in which $\bar{S}S$ and $S\bar{S}$ have codiagonal form. In such a representation, certain linear combinations of $\bar{S}S$ and $S\bar{S}$ can be found with upper and lower triangular form, and these have rational eigenvalues. But the eigenvalues of other linear combinations of X and Y are in general irrational.

We now recapitulate by stating explicitly a simple solution of the classical problem of defining an operator or operators with integral eigenvalues, which can be used to label representations of SO(3) in SU(3).

Let

$$W_1(\nu_1) = (\mu_1 + \nu_1 + 4)\bar{S}S - (\mu_1 + \nu_1)S\bar{S},$$
$$W_2(\nu_2) = (\mu_2 + \nu_2 + 4)S\bar{S} - (\mu_2 + \nu_2)\bar{S}S, \tag{59}$$

where ν_1 and ν_2 are operators whose eigenvalues are the parameters r and s in the states defined by (47), which are related by

$$r + s = \Lambda + \Lambda'. \tag{60}$$

Then it follows directly from (47) that if

$$w_1(\nu_1) = (\mu_1 + \nu_1 + 4)g(\nu_1 - 1)g(\nu_2 + 1)$$
$$- (\mu_1 + \nu_1)g(\nu_1 + 1)g(\nu_2 - 1) + 4(\mu_1 + \nu_1)\{(\mu_2 + 1)[(\mu_2 + 2)^2$$
$$- (\nu_2 + 2)^2] - (\nu_1 + 1)(\mu_2^2 - \nu_2^2)\},$$

$$w_2(\nu_2) = (\mu_2 + \nu_2 + 4)g(\nu_1 + 1)g(\nu_2 - 1) - (\mu_2 + \nu_2)g(\nu_1 - 1)$$
$$\times g(\nu_2 + 1) + 4(\mu_2 + \nu_2)\{(\mu_1 + 1)[(\mu_1 + 2)^2$$
$$- (\nu_1 + 2)^2] - (\nu_2 + 1)(\mu_1^2 - \nu_1^2)\},$$

then

$$[W_1(r) - w_1(r)] | r \rangle$$
$$= 4(\mu_2 - s)(\mu_1 + \mu_2 + r + s + 4)g(r - 1) | r - 2 \rangle,$$
$$[W_2(s) - w_2(s)] | r \rangle = 4(\mu_1 - r)(\mu_1 + \mu_2 + r + s + 4)g(r - 1) | r + 2 \rangle \tag{61}$$

It follows that $w_1(r)$ is an eigenvalue of $W_1(r)$ in a nonorthogonal basis, and that $w_2(r)$ is an eigenvalue of $W_2(r)$ in a different nonorthogonal basis, for each value of r between $\min(\eta, \Lambda + \Lambda' - \mu_2)$ and $\max(\mu_1, \Lambda + \Lambda' - \eta)$. Hence if we define ν_1 and ν_2 by means of the algebraic equations

$$W_1(\nu_1) = w_1(\nu_1), \quad W_2(\nu_2) = w_2(\nu_2), \tag{62}$$

ν_1 and ν_2 will have integral eigenvalues. This may be compared with the definition of l by means of the algebraic equation (51). Either of the operators ν_1 and ν_2 defined by (62) may be used as a labeling operator for the representation of SO(3) in SU(3).

APPENDIX

We now provide a proof, in the notation of this paper, of the results stated in Eqs. (22), (24), (27), and (28).

First we note that, since

$$(\alpha + \tfrac{1}{2}N)^\lambda_\mu = a_{\mu i} a^\lambda_i, \quad (b + \tfrac{1}{2}M)_{ij} = a_{\lambda j} a^\lambda_i, \tag{A1}$$

the relation

$$a_{\lambda j}[(\alpha + \tfrac{1}{2}N)^n]^\lambda_\mu = a_{\mu i}[(b + \tfrac{1}{2}M)^n]_{ij} \tag{A2}$$

is trivially satisfied for $n = 0$ or 1. Also, since α^λ_μ and b_{ij} commute, it follows from (A2) that

$$a_{\lambda j}[(\alpha + \tfrac{1}{2}N)^{n+1}]^\lambda_\mu = a_{\nu i}(a + \tfrac{1}{2}N)^\nu_\mu[(b + \tfrac{1}{2}M)^n]_{ij}$$
$$= a_{\mu k}(b + \tfrac{1}{2}M)_{kl}[(b + \tfrac{1}{2}M)^n]_{ij};$$

so (A2) is true for $n = 2, 3, \cdots$, by induction. If we multiply (A2) on the left by a^μ_j, we obtain the first re-

1381   J. Math. Phys., Vol. 17, No. 8, August 1976

Green, Hurst, and Ilamed   1381

lation of (22). The second relation of (22) is obtained in a similar way by multiplying

$$a^\lambda_j[(\overline{\alpha} - \tfrac{1}{2}N)^n]^\lambda_\mu = a^\mu_i[(\overline{b} - \tfrac{1}{2}M)^n]_{ij} \qquad (A3)$$

on the left by $a_{\mu j}$. Results of this type have been already obtained by the Aguilera-Navarros.[12]

To prove (24), we proceed in a similar way. We note that

$$(S + \tfrac{1}{2}N)^P_Q = a_{Qi}a^P_i, \quad (l + M)_{ij} = a_{Pi}a^P_i, \qquad (A4)$$

so that

$$a_{Qi}[(l + M + 1)^n]_{ij} = a_{Pj}[(S + \tfrac{1}{2}N)^n]^P_Q \qquad (A5)$$

is easily verified for $n = 0$ or 1. Since l_{ij} and S^P_Q commute, it follows by induction that (A5) is true also for $n = 2, 3, \cdots$. The desired relation (24) is obtained from (A5) by multiplying with a^Q_j on the left.

Finally we prove the results (27) and (28) by a similar method. We use the representations

$$A^{\rho\sigma}_{\lambda\mu} = \alpha^\sigma_\mu \delta^\rho_\lambda + \alpha^\sigma_\lambda \delta^\rho_\mu, \quad B_{ijkl} = b_{ik}\delta_{jl} + b_{jk}\delta_{il} \qquad (A6)$$

for the linear operators A and B, and note that the relation

$$a_{\mu k}a_{\lambda l}[(B + M)^n]_{klij} = a_{\rho j}a_{\sigma i}[(A + N)^n]^{\rho\sigma}_{\lambda\mu} \qquad (A7)$$

follows immediately from (A1) and (A6), for $n = 0$ or 1. As B_{klij} and $A^{\rho\sigma}_{\lambda\mu}$ commute, this result follows also by induction for $n = 2, 3, \circ \circ \circ$. If we set $i = j$ in (A7) and multiply on the left by $\alpha^{\lambda\mu}$, the result (27) follows, with the help of

$$\alpha^{\lambda\mu}a_{\mu k}a_{\lambda l} = [(\overline{b} - \tfrac{1}{2}M - 1)(b - \tfrac{1}{2}M)], \text{ etc.} \qquad (A8)$$

In a similar way, we use the representations

$$\overline{A}^{\rho\sigma}_{\lambda\mu} = \overline{\alpha}^\sigma_\mu \delta^\rho_\lambda + \overline{\alpha}^\sigma_\lambda \delta^\rho_\mu, \quad \overline{B}_{ijkl} = b_{ki}\delta_{jl} + \overline{b}_{li}\delta_{jk} \qquad (A9)$$

for $\overline{A}$ and $\overline{B}$ and, establish the relation

$$a^\mu_k a^\lambda_l[(\overline{B} - M)^n]_{ijkl} = a^\rho_j a^\sigma_i[(\overline{A} - N)^n]^{\rho\sigma}_{\lambda\mu}. \qquad (A10)$$

Then we set $i = j$ and multiply this relation on the left by $\alpha_{\lambda\mu}$ to obtain the result (28), with the help of

$$\alpha_{\lambda\mu}a^\mu_k a^\lambda_l = [(b + \tfrac{1}{2}M + 1)(b + \tfrac{1}{2}M)]_{lk}. \qquad (A11)$$

A wide variety of interesting and useful identities can be established by the use of this technique.

*Permanent address: Soreq Institute, Tel Aviv, Israel.

[1]G. Racah, lecture on Lie Groups, in *Lectures of the Istanbul School of Theoretical Physics,* 1962 (Gordon & Breach, New York, 1964).

[2]Y. L. llamed, in *Spectroscopic and Group Theoretical Methods in Physics* (North-Holland, Amsterdam, 1968) pp. 125—130.

[3]J. W. B. Hughes, J. Phys. A **6**, 48, 281, 445, 453 (1973) and Queen Mary College preprint (1974).

[4]B. B. Judd, W. Miller, J. Patera, and P. Winternitz, Univ. of Montreal preprint (1974).

[5]H. S. Green and A. J. Bracken, Int. J. Theor. Phys. **11**, 157 (1974).

[6]V. Bargmann and M. Moshinsky, Nucl. Phys. **18**, 697 (1960); **23**, 177 (1961).

[7]P. Jarvis, J. Phys. A **7**, 1840—16 (1974).

[8]C. A. Hurst and M. A. Lohe, J. Math. Phys. **12**, 1882 (1971).

[9]A. B. Govorkov, Int. J. Theor. Phys. **7**, 49 (1973).

[10]M. Moshinsky and A. C. Quesne, J. Math. Phys. **11**, 1631 (1970); also Phys. Lett. B **29**, 482 (1969).

[11]H. S. Green, J. Math. Phys. **12**, 2106 (1971); see also A. J. Bracken and H. S. Green, J. Math. Phys. **12**, 2099 (1971).

[12]M. C. K. Aguilera-Navarro and V. C. Aguilera-Navarro, J. Math. Phys. **16**, 72 (1974).

[13]A. J. Bracken and H. S. Green, Nuov. Cimento A **9**, 349 (1972).

Particle permutation symmetry of multishell states. I. Two shells*

S. A. Williams and D. L. Pursey

Ames Laboratory-ERDA and Department of Physics, Iowa State University, Ames, Iowa 50011
(Received 20 January 1976)

A method is developed for constructing N-particle states of definite symmetry from n_1-particle and n_2-particle states of definite symmetry where $N = n_1 + n_2$. A canonical resolution of the attendant multiplicity question is given. The results, which are a first step toward the construction of appropriate coefficients of fractional parentage, do not rely upon any particular form for the N-particle Hamiltonian. Rather, the results are based entirely upon properties of the symmetric groups S_{n_1}, S_{n_2}, and S_N. The group theoretic problem which is the construction of irreducible representations of S_N from those of $S_{n_1} \times S_{n_2}$ is solved using induced representation theory together with projection operator techniques.

INTRODUCTION

Among the central mathematical problems in the study of many-particle systems is the construction of an appropriate set of basis states possessing the proper symmetry under the exchange of particles. In the many-nucleon case such basis functions must, of course, be totally antisymmetric, and it is often convenient, as well as physically significant, to accomplish this by using spatial states of some symmetry coupled to spin—isospin states having the conjugate symmetry. In many-particle calculations in which all physical quantities are represented by either one- or two-particle operators, one does not actually need these many-particle states. Rather, one needs the single particle basis states from which they are constructed and the appropriate one- and two-particle coefficients of fractional parentage. These coefficients may be determined once one knows the following: (1) the manner in which the many-particle states are built from the single-particle basis and (2) a complete labelling scheme for the many-particle states including the resolution of any multiplicites.

Formal solutions to the problem of overall antisymmetric states have been known for a very long time,[1] but detailed solutions which are usable in actual calculations have been by and large limited to the following: (1) the case where the only single-particle states involved are degenerate both in energy and angular momentum (i.e., the case of several particles occupying a single shell),[2] (2) special cases involving limited sets of single particle states all degenerate in energy,[3] and (3) the case where the single-particle states are those of the isotropic harmonic oscillator potential.[4-6] The latter has been the most extensively studied and much use is made of the particular symmetries of the oscillator potential to simplify various calculations. The techniques therefore are not very readily generalized to the use of single-particle states arising from a common potential which is somewhat closer to reality. In this paper we present a technique which is independent of the particular potential giving rise to the single-particle basis states, but rather depends for its validity solely upon the properties of the permutation group.

The mathematical problem is that of constructing the basis states of the irreducible representations of $S_{n_1+n_2}$ from those of $S_{n_1} \times S_{n_2}$. That is, we shall construct states of definite permutation symmetry for $n_1 + n_2$ objects from states possessing definite permutation symmetry separately on the first n_1 objects $(1, 2, 3, \cdots, n_1)$ and on the second n_2 objects $(n_1 + 1, n_1 + 2, \cdots, n_1 + n_2)$. This is accomplished by induced representation theory applied to the symmetric group and by making use of projection operations. The results are applicable whatever the nature of the objects.

As an example for motivation one might consider the first n_1 objects to be nucleons $(1, 2, 3, \cdots, n_1)$ which belong to some shell appropriate to a common but unspecified potential. The second n_2 nucleons $(n_1 + 1, n_1 + 2, \cdots, n_1 + n_2)$ then belong to some other shell. By a shell, one would mean a set of single-particle states of this common potential which are degenerate in energy and share the same angular momentum. This angular momentum could be orbital only, in which case one would need the appropriate spin—isospin states to produce overall antisymmetry. The spin—isospin problem has essentially been solved.[7,8] If the angular momentum is total (orbital plus spin), then appropriate isospin functions are required; the solution of this problem is also well known.[9] The coupling of the individual angular momenta of particles within a shell to states labelled by the shell angular momentum and the coupling of the shell angular momenta to an $(n_1 + n_2)$-particle state of definite angular momentum is a separate problem which may be handled by standard angular momentum techniques.[9] To make use of our formalism it is necessary to regard the states of the n_1 particles in the first shell as having been arranged to form the basis functions for some irreducible representation of the symmetric group S_{n_1}; similarly n_2 and S_{n_2}. This latter problem, which is not entirely divorced from the question of shell angular momentum, is solved in principle and in practice by the use of standard techniques.[8-10] It remains unsolved, however, in any truly elegant manner.

In a future paper we intend to generalize the formalism to the case of $S_{n_1+n_2+n_3 \cdots + n_k}$ basis functions expressed in terms of those of $S_{n_1} \times S_{n_2} \times S_{n_3} \times \cdots \times S_{n_k}$. The techniques used in this first paper are, we feel, close to the experience of most many-particle theorists. In future publications we intend to also deal with the problem of coefficients of fractional parentage and with the question of spurious excitations of the center of mass which are inherent in any many-particle, common potential basis.

In Sec. I a simple sketch of the basic ideas of induced representations is presented as this is central to the entire approach. Also, the example of constructing states for S_3 from those of $S_1 \times S_2$ by "brute force" methods is presented. In Sec. II the general problem of constructing states of $S_{n_1+n_2}$ from those of $S_{n_1} \times S_{n_2}$ is presented and our main results are in Sec. III. Also in Sec. III we consider the multiplicity and orthogonality problems and illustrate our formal results by once again returning to the S_3 example. Section IV is a final example and commentary.

I. INDUCED REPRESENTATIONS AND AN EXAMPLE

The basic problem is, given the direct product states

$$| n_1 \lambda_1 r_1; n_2 \lambda_2 r_2 \rangle \tag{1}$$

how does one construct the states

$$| n_1 + n_2 \; \lambda r; n_1 \lambda_1, n_2 \lambda_2 \rangle \; ? \tag{2}$$

In (1) the labeling means the state belongs to the r_1 row of the λ_1 irreducible representation (IR) of S_{n_1} and to the r_2 row of the λ_2 IR of S_{n_2}. It is to be understood that the permutations of S_{n_1} are permutations of the first n_1 objects $(1, 2, \cdots, n_1)$ while those of S_{n_2} are permutations of the second n_2 objects $(n_1 + 1, n_1 + 2, \cdots, n_1 + n_2)$. The set of all states (1) labeled r_1 and r_2 for fixed λ_1 and λ_2 are the basis for the direct product representation $\lambda_1 \times \lambda_2$ of $S_{n_1+n_2}$. The states (2) belong to the r row of the λ IR of $S_{n_1+n_2}$, and the extra labels are to remind us how this state was constructed. The product representation $\lambda_1 \times \lambda_2$ which is an IR of $S_{n_1} \times S_{n_2}$ induces a representation of $S_{n_1+n_2}$ which is to be broken into irreducible parts. The state $| n_1 + n_2 \; \lambda r; n_1 \lambda_1, n_2 \lambda_2 \rangle$ will be a linear combination of all the states r_1, r_2 of the form (1) together with other states having the same form as (1) but refering to $S'_{n_1} \times S'_{n_2}$ in which the permutations of S'_{n_1} involve some different n_1 of the $n_1 + n_2$ objects and those of S'_{n_2} involve the remaining n_2 objects. In fact, the state $| n_1 + n_2 \; \lambda r; n_1 \lambda_1, n_2 \lambda_2 \rangle$ will be a linear combination of states involving all possible ways of selecting n_1 objects out of the total $n_1 + n_2$ objects. From now on we shall omit the labels n_1, n_2 and $n_1 + n_2$ from the basis vectors.

We consider first of all the simple example in which we have one particle (1) in one shell and two other particles (2 and 3) in a different shell. The particle 1 must be described by a state which belongs to the $[1]$ IR of S_1.[11] For S_1 this is the only IR and it is one dimensional. The state of particles 2 and 3 may be either symmetric under the exchange of 2 and 3 or antisymmetric. In the former case it belongs to the $[2]$ IR of S_2 while in the latter it belongs to the $[1, 1]$ IR. For this example we choose to use the $[2]$ IR. There is then only one direct product basis state $\varphi(1; 2, 3)$ whence

$$\varphi(1; 2, 3) = U(1) V(2, 3) \tag{3}$$

in which $U(1)$ is the state of particle 1 belonging to the $[1]$ IR of S_1 and $V(2, 3)$ is the symmetric state of particles 2 and 3 and belongs to the $[2]$ IR of S_2.

The possible states of three particles may be grouped into basis functions of the IR of S_3. There are three such IR: $[3]$, $[2, 1]$, and $[1, 1, 1]$. The first problem is to determine which of these is compatible with the state Eq.

(3). By standard tableaux multiplication one has

$$[1] \times [2] = [3] + [21]. \tag{4}$$

The reducible representation of S_3 on the right of Eq. (4) is essentially the representation of S_3 induced from the IR $[1] \times [2]$ of $S_1 \times S_2$. The next problem is to express the basis states for either the $[3]$ or $[21]$ IR contained in the induced representation of S_3 in terms of states like that of Eq. (3). In particular we will need the states in which each in turn of 1, 2, or 3 is in state U while the remaining pair is in state V. We shall use the shorthand notation

$$\varphi_1 \equiv \varphi(1; 2, 3) = U(1) V(2, 3),$$

$$\varphi_2 \equiv \varphi(2; 3, 1) = U(2) V(3, 1), \tag{5}$$

$$\varphi_3 \equiv \varphi(3; 1, 2) = U(3) V(1, 2)$$

and regard these as a column vector. Each of these is the basis for the $[1] \times [2]$ IR of $S_1 \times S_2$, but they differ in the identification of the particular $S_1 \times S_2$ subgroups of S_3.

In the basis given by Eqs. (5), the matrix representatives of the permutations (12) and (13) are

$$(12) = \begin{pmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \end{pmatrix}, \quad (13) = \begin{pmatrix} 0 & 0 & 1 \\ 0 & 1 & 0 \\ 1 & 0 & 0 \end{pmatrix}. \tag{6}$$

All the permutations of S_3 may be built from products of these two.

We next order the basis function of the $[3]$ and $[21]$ IR of S_3 contained in the reduced representation as

$$\psi_1 \equiv \psi([3] \; (111); \; [1], [2]),$$

$$\psi_2 \equiv \psi([21] \; (211); \; [1], [2]), \tag{7}$$

$$\psi_3 \equiv \psi([21] \; (121); \; [1], [2].$$

In Eq. (7) the essential labels are the $[3]$ and $[21]$ which label the IR of S_3 and the (111), (211), and (121) which are Yamanouchi symbols labeling the rows within an IR. In the basis of Eq. (7) the permutations (12) and (13) have the standard forms (indicated by superscripts)

$$(12)^s = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & -1 \end{pmatrix}, \quad (13)^s = \begin{pmatrix} 1 & 0 & 0 \\ 0 & -1/2 & -\sqrt{3}/2 \\ 0 & -\sqrt{3}/2 & 1/2 \end{pmatrix}. \tag{8}$$

Equations (6) express the induced representation in terms of the basis [Eq. (5)] which is constructed naturally from the IR $[1] \times [2]$ of $S_1 \times S_2$. On the other hand, Eqs. (8) express the same induced representation in terms of the basis [Eq. (7)] which displays the irreducible components of the representation in an explicit manner.

The problem at hand then is to determine the transformation which expresses the basis of Eq. (7) in terms of the basis of Eq. (5). This transformation is unitary and is defined by

$$\psi_j = \sum_{k=1}^{3} U_{jk} \varphi_k. \tag{9}$$

For any permutation of S_3 represented by π^s in the standard basis, Eq. (7), and by π in the natural induced basis, Eq. (5), we must have

$$\pi^s U = U\pi, \qquad (10)$$

in which U is the matrix whose elements are the U_{jk} of Eq. (9). By using standard matrix techniques together with Eqs. (6) and (8) we find

$$U = \begin{pmatrix} 1/\sqrt{3} & 1/\sqrt{3} & 1/\sqrt{3} \\ \sqrt{6}/6 & \sqrt{6}/6 & -\sqrt{6}/3 \\ 1/\sqrt{2} & 1/\sqrt{2} & 0 \end{pmatrix}. \qquad (11)$$

From Eqs. (11) and (9) we then have, for example,

$$\psi([21]\ (211);\ [1][2]) = (\sqrt{6}/6)\varphi(1;\ 2,\ 3) + (\sqrt{6}/6)\varphi(2;\ 3,\ 1)$$
$$- (\sqrt{6}/3)\varphi(3;\ 1,\ 2). \qquad (12)$$

Now, we want to examine what we have accomplished by "brute force." First of all we note that the subgroup $S_1 \times S_2$ of S_3 contains the elements e and (23). The left cosets of $S_1 \times S_2$ in S_3 are the sets of elements $\{\pi\pi'\}$, where π is any fixed element of S_3 and different elements of the left coset defined by π are obtained as π' runs over the elements of the subgroup $S_1 \times S_2$. Any element of a left coset can serve to label the coset, and we may arbitrarily choose a standard labeling element for each such coset. These labeling elements are called coset representatives. From the group multiplication table of S_3, Table I, we may construct the entries in Table II, which indicates the left coset to which each element of S_3 belongs together with a coset representative. The basis states φ_1, φ_2, φ_3 [Eq. (5)] in terms of which we expand the ψ_1, ψ_2, ψ_3 [Eq. (1)] may all be formed from $\varphi(1;\ 2,\ 3)$ by

$$\varphi_1 \equiv \varphi(1;\ 2,\ 3) = e\varphi(1;\ 2,\ 3) \equiv \varphi_e,$$

$$\varphi_2 \equiv \varphi(2;\ 3,\ 1) = (12)\varphi(1;\ 2,\ 3) \equiv \varphi_{(12)}, \qquad (13)$$

$$\varphi_3 \equiv \varphi(3;\ 1,\ 2) = (13)\varphi(1;\ 2,\ 3) \equiv \varphi_{(13)}.$$

Hence, the basis states of Eq. (5) are formed from the basis states of the IR $[1] \times [2]$ of the subgroup $S_1 \times S_2$ [in this case the single state $\varphi(1;\ 2,\ 3)$] by operating with coset representatives of the left cosets of $S_1 \times S_2$ in S_3. The dimension of the induced representation of S_3 is clearly $d([1])\ d([2]) \times$ the number of left cosets of $S_1 \times S_2$ in S_3, where $d(\lambda)$ is the dimension of the $[\lambda]$ IR. In this example the induced representation has dimension three and decomposes into the one-dimensional IR, $[3]$, plus the two-dimensional IR, $[21]$, of S_3. As indicated in Eqs. (13) the natural basis for the induced representation may be labeled by the label $[1] \times [2]$ of the IR of $S_1 \times S_2$, the row labels within this IR, and the left coset representation; in this example we have omitted the first two of these labels.

We now give a more general review of those aspects of induced representation theory which are central to the remainder of the paper. Suppose H is a subgroup of order (H : 1) of some group G of order (G : 1). Each element g of G belongs to one and only one left coset of H in G and may be written as

$$g = g_j h(g), \qquad (14)$$

where g_j is the representative of the appropriate left coset and $h(g)$ is an element of H depending upon g. The number of left cosets is the order of G divided by the order of H, i.e.,

$$(G:1)/(H:1) \equiv (G:H). \qquad (15)$$

Now, let $|\ \mu l\rangle$ be a basis state for the lth row of the $[\mu]$ IR of H; the dimension of $[\mu]$ is $d(\mu)$. The set of states

$$g_j |\ \mu l\rangle \equiv |\ \mu l;\ g_j\rangle,$$

$$l = 1,\ 2 \cdots d(\mu), \quad j = 1,\ 2 \cdots (G:H),$$

forms a natural basis for the representation of G induced by $[\mu]$ of H; the dimension of this induced representation is $(G:H)\ d(\mu)$. From Eq. (14) it follows that

$$g |\ \mu l;\ g_j\rangle = g' |\ \mu l\rangle,$$

where $g' = gg_j$ belongs to G. Hence also g' belongs to the some (unique) left coset of H in G, say that labelled by g_m. So $g' = g_m h(g')$. Since $|\ \mu l\rangle$ belongs to the $[\mu]$ IR of H, and since $h(g')$ belongs to H,

$$g |\ \mu l;\ g_j\rangle = g_m h(g') |\ \mu l\rangle$$
$$= \sum_{l'} D^\mu_{l'l}(h(g')) |\ \mu l';\ g_m\rangle, \qquad (16)$$

where $D^\mu_{l'l}$ is the matrix element $\langle \mu l' | h(g') | \mu l\rangle$ of $h(g')$ in the $[\mu]$ IR. We may make Eq. (16) more transparent by defining

$$\widetilde{D}^\mu_{l'l}(g \in G) = \begin{cases} D^\mu_{l'l}(g) & \text{if } g \in H, \\ 0 & \text{if } g \notin H. \end{cases} \qquad (17)$$

In Eq. (16) then one may sum over left coset representatives and the $h(g') = g_m^{-1} g g_j$ will pick out the appropriate one. Thus,

$$g |\ \mu l;\ g_j\rangle = \sum_{l'g_{j'}} \widetilde{D}^\mu_{l'l}(g_{j'}^{-1} g g_j) |\ \mu l';\ g_{j'}\rangle. \qquad (18)$$

Hence the matrix of the representation $[\mu \uparrow G]$ of G induced by the $[\mu]$ IR of H is given by

$$D^{\mu \uparrow G}_{l'j',lj}(g) = \widetilde{D}^\mu_{l'l}(g_{j'}^{-1} g g_j). \qquad (19)$$

The general problem then is to determine which IR of G occur in this induced representation, how often each IR occurs (the multiplicity question), and specifically how to determine appropriate bases to reduce the induced representation into irreducible parts. This was

TABLE I. Group multiplication table for S_3.

e	(12)	(13)	(23)	(123)	(132)
(12)	e	(132)	(123)	(23)	(13)
(13)	(123)	e	(132)	(12)	(23)
(23)	(132)	(123)	e	(13)	(12)
(123)	(13)	(23)	(12)	(132)	e
(132)	(23)	(12)	(13)	e	(123)

TABLE II. Left coset decomposition of S_3 with respect to $S_1 \times S_2$.

$\pi \in S_3$	left coset	coset representative
e	$[e, (23)]$	e
(12)	$[(12), (123)]$	(12)
(13)	$[(13), (132)]$	(13)
(23)	$[(23), e]$	e
(123)	$[(23), (12)]$	(12)
(132)	$[(132), (13)]$	(13)

done for S_3 by inspection. In the more general case we shall use projection operators often called the method of idempotents.

Consider the operator

$$P^\lambda_{jk} \equiv \frac{d(\lambda)}{(G:1)} \sum_{g \in G} D^{\lambda*}_{jk}(g) g, \tag{20}$$

where $[\lambda]$ is the $d(\lambda)$-dimensional IR of G which we have presumed to be unitary; the sum is over all elements $g \in G$. The Hermitian conjugate operator is

$$P^{\lambda\dagger}_{jk} = \frac{d(\lambda)}{(G:1)} \sum_{g \in G} D^{\lambda}_{jk}(g) g^{-1} = P^\lambda_{kj} \tag{21}$$

so that

$$P^{\lambda\dagger}_{jk} P^{\lambda'}_{j'k'} = \delta_{jj'} \delta_{\lambda\lambda'} P^\lambda_{kk'}. \tag{22}$$

follows from Eq. (20) together with the representation property of the matrices D^λ.

Now, we define a state belonging to the jth row of the λth IR of G projected from the state $|\mu l; g_m\rangle$ to be

$$|\lambda j : k\mu l; g_m\rangle \equiv P^\lambda_{jk} |\mu l; g_m\rangle. \tag{23}$$

In Eq. (23) the extra labels $k\mu l$ and g_m serve as potential multiplicity labels. By inserting Eq. (20) into Eq. (23) one finds

$$|\lambda j : k\mu l; g_m\rangle = \sum_{l' g_{m'}} \langle \mu l'; g_{m'} | \lambda j : k\mu l; g_m\rangle |\mu l'; g_{m'}\rangle, \tag{24}$$

where the transformation bracket is given by

$$\langle \mu l'; g_{m'} | \lambda j : k\mu l; g_m\rangle = \frac{d(\lambda)}{(G:1)} \sum_{g \in G} D^{\lambda*}_{jk}(g) \tilde{D}^\mu_{l',l}(g^{-1}_{m'} g g_m). \tag{25}$$

From Eq. (23) and the orthogonality of the bases $|\mu l; g_m\rangle$ it follows also that

$$\langle \mu l'; g_{m'} | \lambda j : k\mu l; g_m\rangle = \langle \mu l'; g_{m'} | P^\lambda_{jk} | \mu l; g_m\rangle. \tag{26}$$

The projected states are orthogonal on λ and j but not upon the multiplicity labels. The overlap of two such states follows from Eqs. (22) and (23) and is

$$\begin{aligned}
\langle \lambda' j' : k'\mu l'; g_{m'} | \lambda j : k\mu l; g_m\rangle &= \langle \mu l'; g_{m'} | P^{\lambda'\dagger}_{j'k'} P^\lambda_{jk} | \mu l; g_m\rangle \\
&= \delta_{j'j} \delta_{\lambda'\lambda} \langle \mu l'; g_{m'} | P^\lambda_{k'k} | \mu l; g_m\rangle \\
&= \delta_{j'j} \delta_{\lambda'\lambda} \langle \mu l'; g_{m'} | \lambda k' ; k\mu l; g_m\rangle.
\end{aligned} \tag{27}$$

Hence, for states belonging to the same row of the same IR of G, the overlap matrix is dependent only upon the multiplicity labels.

Equation (25) which gives the transformation brackets and also the overlap matrix elements may be processed further. From Eq. (17) it follows that in summing over $g \in G$ there is no contribution unless $g^{-1}_m g g_m = h \in H$. Hence, we may replace the sum on g by a sum over $h \in H$ and write $g = g_m h g^{-1}_m$. Thus

$$\begin{aligned}
\langle \mu l'; g_{m'} | \lambda j : k\mu l; g_m\rangle &= \frac{d(\lambda)}{(G:1)} \sum_{h \in H} D^{\lambda*}_{jk}(g_m h g^{-1}_m) D^\mu_{l',l}(h) \\
&= \frac{d(\lambda)}{(G:1)} \sum_{\substack{h \in H \\ st}} D^{\lambda*}_{js}(g_{m'}) D^{\lambda*}_{tk}(g^{-1}_m) D^{\lambda*}_{st}(h)
\end{aligned}$$

We may write Eq. (28) in a very useful form by noting that the operator which projects onto the l state of the $[\mu]$ IR of H is

$$\begin{aligned}
\rho^\mu_{ll'} &= \frac{d(\mu)}{(H:1)} \sum_h D^{\mu*}_{ll'}(h) h = \frac{d(\mu)}{(H:1)} \sum_h D^{\mu*}_{ll'}(h^{-1}) h^{-1} \\
&= \frac{d(\mu)}{(H:1)} \sum_h D^\mu_{l',l}(h) h^{-1}.
\end{aligned} \tag{29}$$

so that

$$\langle \lambda t | \rho^\mu_{ll'} | \lambda s\rangle = \frac{d(\mu)}{(H:1)} \sum_{h \in H} D^{\lambda*}_{st}(h) D^\mu_{l',l}(h). \tag{30}$$

Thus,

$$\begin{aligned}
\langle \mu l'; g_{m'} | \lambda j : k\mu l; g_m\rangle &= \frac{d(\lambda)}{(G:1)} \frac{(H:1)}{d(\mu)} \sum_{st} D^{\lambda*}_{js}(g_{m'}) D^{\lambda*}_{tk}(g^{-1}_m) \\
&\quad \times \langle \lambda t | \rho^\mu_{ll'} | \lambda s\rangle \\
&= \frac{d(\lambda)}{(G:1)} \frac{(H:1)}{d(\mu)} \langle \lambda k | g_m \rho^\mu_{ll'} g^{-1}_{m'} | \lambda j\rangle.
\end{aligned} \tag{31}$$

The equality of Eqs. (27) and (31) is a manifestation of the reciprocity theorems of Froebenius.[12] Specifically, if the $[\lambda]$ IR of G is induced from the $[\mu]$ IR of H $n(\lambda, \mu)$ times and the $[\mu]$ IR of H is subduced from the $[\lambda]$ IR of G $n(\mu, \lambda)$ times, then $n(\lambda, \mu) = n(\mu, \lambda)$. Thus the multiplicity is given by

$$n(\mu, \lambda) = n(\lambda, \mu) = \frac{1}{(H:1)} \sum_{h \in H} \chi^{\lambda*}(h) \chi^\mu(h), \tag{32}$$

in which $\chi^\lambda(h) = \sum_j D^\lambda_{jj}(h)$ is the character of h in the $[\lambda]$ IR of G and $\chi^\mu(h) = \sum_l D^\mu_{ll}(h)$ is the character of h in the $[\mu]$ IR of H.

We have introduced three labels, k, l, and g_m, to distinguish the (possibly degenerate) states of the IR $[\lambda]$. In fact, all the occurrences of $[\lambda]$ may be projected from the states $|\mu l; g_m\rangle$ with arbitrarily chosen but fixed values of l and g_m. To see this, we use the results (i) for every group element g of G,

$$P^\lambda_{jk} g = \sum_{k'} P^\lambda_{jk'} D^\lambda_{k'k}(g), \tag{33}$$

(ii)

$$|\mu l; g_m\rangle = g_m | \mu l; e\rangle,$$

and (iii)

$$|\mu l'; e\rangle = \rho^\mu_{l',l} | \mu l; e\rangle.$$

Hence

$$\begin{aligned}
|\lambda j : k\mu l'; g_{m'}\rangle &= P^\lambda_{jk} | \mu l; g_{m'}\rangle \\
&= P^\lambda_{jk} (g_{m'} \rho^\mu_{l',l} g^{-1}_{m'}) | \mu l; g_{m'}\rangle \\
&= \sum_{k'} c_{k'} P^\lambda_{jk'} | \mu l; g_{m'}\rangle \\
&= \sum_{k'} c_{k'} | \lambda j : k'\mu l ; g_{m'}\rangle,
\end{aligned} \tag{34a}$$

where

$$c_{k'} = \frac{d(\mu)}{(H:1)} \sum_{h \in H} D^\mu_{l',l}(h)^* D^\lambda_{kk'}(g_{m'} h g^{-1}_{m'}). \tag{34b}$$

Hence the state $|\lambda j : k\mu l'; g_{m'}\rangle$ has been expressed as

a linear combination of the states $|\lambda j : k' \mu l; g_m\rangle$ with arbitrarily chosen but fixed l and g_m. It is most convenient to choose $g_m = e$, as was done in the S_3 example, and we shall make this choice in all that follows. We are left with $d(\lambda)$ possible values of the remaining degeneracy label k. Of course, these need not yield distinct states.

It is now convenient to relate the number of occurrences of $[\lambda]$ in $[\mu \uparrow G]$, viz., $n(\lambda, \mu)$ to the trace of the $d(\lambda) \times d(\lambda)$ overlap matrix $\langle \lambda j : k\mu l; e | \lambda j : k' \mu l; e \rangle$ (for arbitrary but fixed l). We have, from Eqs. (27) and (31),

$$\sum_k \langle \lambda j : k\mu l; e | \lambda j : k\mu l; e \rangle$$

$$= \frac{d(\lambda)}{(G:1)} \frac{(H:1)}{d(\mu)} \sum_k \langle \lambda k | P^\mu_{ll} | \lambda k \rangle$$

$$= \frac{d(\lambda)}{(G:1)} \sum_{h \in H} \sum_k \langle \lambda k | h | \lambda k \rangle D^\mu_{ll}(h)^*$$

$$= \frac{d(\lambda)}{(G:1)} \sum_{h \in H} \chi^\lambda(h) D^\mu_{ll}(h)^*.$$

But from Eqs. (34) this expression can be shown to be independent of l. Hence we may average over l to obtain

$$\sum_k \langle \lambda j : k\mu l; e | \lambda j : k\mu l; e \rangle$$

$$= \frac{d(\lambda)}{(G:1)} \frac{1}{d(\mu)} \sum_{h \in H} \chi^\lambda(h) \chi^\mu(h)^*$$

$$= \frac{d(\lambda)}{(G:1)} \frac{(H:1)}{d(\mu)} n(\lambda, \mu). \tag{35}$$

For fixed l and $g_m = e$, we adopt the simplified notation

$$|\lambda j : k\mu l; e\rangle \equiv |\lambda j : k\mu l\rangle$$

and have

$$|\lambda j : k\mu l\rangle = P^\lambda_{jk} |\mu l; e\rangle = \sum_{l' g_{m'}} \langle \mu l'; g_m | \lambda j : k\mu l\rangle |\mu l'; g_{m'}\rangle,$$

$$\tag{36}$$

where

$$\langle \mu l'; g_{m'} | \lambda j : k\mu l\rangle = \frac{d(\lambda)}{(G:1)} \frac{(H:1)}{d(\mu)} \sum_s D^{\lambda *}_{js}(g_{m'}) \langle \lambda k | P^\mu_{ll'} | \lambda s\rangle$$

$$= \frac{d(\lambda)}{(G:1)} \sum_{\substack{h \in H \\ s}} D^{\lambda *}_{js}(g_{m'}) D^\lambda_{ks}(h) D^{\mu *}_{ll'}(h). \tag{37}$$

The overlap of these states is

$$\langle \lambda j : k' \mu l | \lambda j : k\mu l\rangle = \frac{d(\lambda)}{(G:1)} \frac{(H:1)}{d(\mu)} \langle \lambda k | P^\mu_{ll} | \lambda k'\rangle$$

$$= \frac{d(\lambda)}{(G:1)} \sum_{h \in H} D^\lambda_{kk'}(h) D^{\mu *}_{ll}(h). \tag{38}$$

These projected states are not normalized nor are they orthogonal on the multiplicity label k. The normalization factor is the square root of the overlap of the state with itself; we choose the phase convention of taking the positive square root. Thus

$$\langle \lambda j : k\mu l | \lambda j : k\mu l\rangle = \frac{d(\lambda)}{(G:1)} \frac{(H:1)}{d(\mu)} \langle \lambda k | P^\mu_{ll} | \lambda k\rangle$$

$$= \frac{d(\lambda)}{(G:1)} \sum_{h \in H} D^\lambda_{kk}(h) D^{\mu *}_{ll}(h). \tag{39}$$

From Eq. (34), the induced representation is guaranteed to be not more dense than the regular representation of G. That is, $[\lambda]$ occurs at most $d(\lambda)$ times. In general, the multiplicity of $[\lambda]$ will be considerably less, the actual number of occurrences being given by $(G:H)d(\mu)/d(\lambda)$ times the trace of the overlap matrix for the nonnormalized projected states. Thus the possible values of k overdetermine the multiple occurrences of $[\lambda]$ in $[\mu \uparrow G]$. Many of the states projected with different k values will be identical when normalized. One therefore needs criteria by which nonidentical states may be chosen up to $n(\lambda, \mu)$ in number. Such a criterion is afforded by considering normalized projected states

$$|\lambda j : k\mu l) = |\lambda j : k\mu l\rangle / \sqrt{\langle \lambda j : k\mu l | \lambda j : k\mu l\rangle} \tag{40}$$

Thus, two states are identical if

$$\langle \lambda j : k' \mu l | \lambda j : k\mu l\rangle$$

$$= \sqrt{\langle \lambda j : k\mu l | \lambda j : k\mu l\rangle \langle \lambda j : k' \mu l | \lambda j : k' \mu l\rangle} \tag{41a}$$

or

$$\langle \lambda k' | P^\mu_{ll} | \lambda k\rangle = (\langle \lambda k' | P^\mu_{ll} | \lambda k\rangle \langle \lambda k' | P^\mu_{ll} | \lambda k'\rangle)^{1/2}. \tag{41b}$$

We notice in passing that Eq. (41b) is a reciprocity statement again, in that it says that if in the subduced representation of H, the states of $[\mu]$ of H projected from $|\lambda k\rangle$ of the $[\lambda]$ IR of G using P^μ_{jl} are identical with those of $[\mu]$ projected using P^μ_{jl} on $|\lambda k'\rangle$, then the states of $[\lambda]$ projected from $|\mu l\rangle$ of H using P^λ_{jk} are the same as those of $[\lambda]$ projected from $|\mu l\rangle$ using $P^\lambda_{jk'}$. For practical purposes, however, it matters not which expression one uses, for in either case it is the overlap and normalization factors which must be computed.

The remainder of this paper is devoted to exploiting the foregoing method for the case of $S_{n_1+n_2} \supset S_{n_1} \times S_{n_2}$.

II. LEFT COSETS OF $S_{n_1} \times S_{n_2}$ IN $S_{n_1+n_2}$

In this section we shall enumerate the left cosets of $S_{n_1} \times S_{n_2}$ in $S_{n_1+n_2}$ by giving a complete set of left coset representatives. The number of elements of $S_{n_1} \times S_{n_2}$ is $n_1! n_2!$ and the number of elements of $S_{n_1+n_2}$ is $(n_1 + n_2)!$. Thus, the number of left cosets is

$$\binom{n_1 + n_2}{n_1} = \frac{(n_1 + n_2)!}{n_1! n_2!}.$$

Consider the set of transpositions (x, y), where x is any of the symbols $\{1, 2 \cdots n_1\}$ and y is any of $\{n_1 + 1, n_1 + 2 \cdots n_1 + n_2\}$. There are

$$n_1 n_2 = \binom{n_1}{1} \binom{n_2}{1}$$

such transpositions. Next consider the set of products of these of the form $(x_1, y_1)(x_2, y_2)$ with the restriction $x_1 < x_2, y_1 < y_2$; there are

$$\binom{n_1}{2} \binom{n_2}{2}$$

such doubles. In a similar manner there are

$$\binom{n_1}{3}\binom{n_2}{3}$$

triples $(x_1, y_1)(x_2, y_2)(x_3, y_3)$, $x_1 < x_2 < x_3$, $y_1 < y_2 < y_3$. The total number of such elements of $S_{n_1+n_2}$, including the identity, is

$$\sum_{r=0}^{}\binom{n_1}{r}\binom{n_2}{r} = \binom{n_1+n_2}{n_1}.$$

Thus the set

$$\mathcal{S} = \{e, (x, y), (x_1, y_1)(x_2, y_2), \cdots\}$$

$$x_j \in \{1, 2, 3 \cdots n_1\}, \quad y_j \in \{n_1+1, n_1+2, \ldots, n_1+n_2\},$$

$$x_j < x_{j+1}, \quad y_j < y_{j+1}, \tag{42}$$

has precisely as many members as there are left cosets. It remains to show that these are suitable left coset representatives which we shall do by explicitly demonstrating that an arbitrary element π of $S_{n_1+n_2}$ may be written as

$$\pi = g_j \pi_1 \pi_2,$$

where $g_j \in \mathcal{S}$ and $\pi_1 \in S_{n_1}$, $\pi_2 \in S_{n_2}$. One must keep in mind that the elements of S_{n_1} are permutations on the symbols $1, 2 \cdots n_1$ and those of S_{n_2} are permutations on the symbols $n_1+1, n_1+2, \ldots, n_1+n_2$.

A typical element of $S_{n_1+n_2}$ has the form

$$\begin{pmatrix} 1\ 2 \cdots n_1 & \vdots & n_1+1, n_1+2 \cdots n_1+n_2 \\ \cdots y_{j_1} \cdots y_{j_2} \cdots y_{j_3} \cdots & \vdots & \cdots x_{k_1} \cdots x_{k_2} \cdots x_{k_3} \cdots \end{pmatrix},$$

in which $1 \le x_1 < x_2 < x_3 \cdots \le n_1$ and $n_1+1 \le y_1 < y_2 \cdots \le n_1+n_2$. The indices $j_1, j_2 \cdots$ must then be some permutation of the induces $k_1, k_2 \cdots$. We have divided the permutation by a vertical dashed line to indicate the $S_{n_1} \times S_{n_2}$ decomposition. By direct multiplication one readily finds

$$\begin{pmatrix} 1\ 2 \cdots n_1 & \vdots & n_1+1, \cdots n_1+n_2 \\ \cdots y_{j_1} \cdots y_{j_2} \cdots y_{j_3} & \vdots & \cdots x_{k_1} \cdots x_{k_2} \cdots x_{k_3} \cdots \end{pmatrix}$$

$$= (x_1, y_1)(x_2, y_2) \cdots$$

$$\times \begin{pmatrix} 1\ 2 \cdots n_1 & \vdots & n_1+1, n_1+2 \cdots n_1+n_2 \\ \circ \cdot x_{j_1} \cdot \cdot x_{j_2} \cdots x_{j_3} \cdots & \vdots & \cdots y_{k_1} \cdots y_{k_2} \cdots y_{k_3} \cdots \end{pmatrix}, \tag{43}$$

which proves the result. The "out of place" symbols are ordered in increasing order on each side of the dashed line and are brought out as transpositions which involve one of $1, 2 \cdots n_1$ together with one of $n_1+1, n_1+2, \cdots$, n_1+n_2. This is done pairwise in increasing subscript order which yields a factor from the set $\mathcal{S}$. The remaining permutation obviously has the form $\pi_1\pi_2$ and is constructed from the original by interchanging x_{j_1} with y_{j_1}, x_{j_2} with y_{j_2}, etc. An example from $S_9 \supset S_3 \times S_6$ is

$$\begin{pmatrix} 1 & 2 & 3 & \vdots & 4 & 5 & 6 & 7 & 8 & 9 \\ 9 & 4 & 6 & \vdots & 1 & 7 & 5 & 8 & 2 & 3 \end{pmatrix}$$
$$\begin{array}{ccccccc} y_3 & y_1 & y_2 & x_1 & & x_2 & x_3 \end{array}$$

$$= (14)(26)(39)\begin{pmatrix} 1 & 2 & 3 & \vdots & 4 & 5 & 6 & 7 & 8 & 9 \\ 3 & 1 & 2 & \vdots & 4 & 7 & 5 & 8 & 6 & 9 \end{pmatrix}$$

$$= \underbrace{(14)(26)(39)}_{g_j \in \mathcal{S}} \quad \underset{\pi_1 \in S_3}{(132)} \quad \underset{\pi_2 \in S_6}{(5786)}$$

Similarly, for $S_9 \supset S_4 \times S_5$ the same permutation is decomposed as

$$\begin{pmatrix} 1 & 2 & 3 & 4 & \vdots & 5 & 6 & 7 & 8 & 9 \\ 9 & 4 & 6 & 1 & \vdots & 7 & 5 & 8 & 2 & 3 \end{pmatrix} = (26)(39)(1324)(5786).$$

These results now allow us to apply the methods of Sec. I to establish a natural basis for the representation of $S_{n_1+n_2}$ induced by the $[\lambda_1] \times [\lambda_2]$ IR of $S_{n_1} \times S_{n_2}$.

III. THE $S_{n_1+n_2} \supset S_{n_1} \times S_{n_2}$ PROBLEMS

The basis states for the representation $[\lambda_1] \times [\lambda_2]$ of $S_{n_1+n_2}$ induced by the $[\lambda_1] \times [\lambda_2]$ IR of $S_{n_1} \times S_{n_2}$ are labeled $|\lambda_1 s_1, \lambda_2 s_2; g_j\rangle$, where $\lambda_1 s_1$ labels the s_1 row of the $[\lambda_1]$ IR of S_{n_1} and $\lambda_2 s_2$ the s_2 row of the $[\lambda_2]$ IR of S_{n_2}; g_j is a left coset representative. A state belonging to the s occurrence of the rth row of the $[\lambda]$ IR of $S_{n_1+n_2}$ is then projected according to Eq. (36) as

$$|\lambda r : s\lambda_1 s_1 \lambda_2 s_2\rangle = P_{rs}^\lambda |\lambda_1 s_1 \lambda_2 s_2\rangle$$

$$= \sum_{g_j t_1 t_2} \langle \lambda_1 t_1 \lambda_2 t_2; g_j | \lambda r : s\lambda_1 s_1 \lambda_2 s_2\rangle$$

$$\times |\lambda_1 t_1 \lambda_2 t_2; g_j\rangle, \tag{44}$$

in which the transposition brackets are given by [after Eq. (37)]

$$\langle \lambda_1 t_1 \lambda_2 t_2; g_j | \lambda r : s\lambda_1 s_1 \lambda_2 s_2\rangle$$

$$= \frac{d(\lambda)}{(n_1+n_2)!} \sum_{\substack{\pi_1\pi_2 \\ t\ v}} D_{rt}^{\lambda*}(g_j) D_{sv}^\lambda(\pi_1) D_{vt}^\lambda(\pi_2)$$

$$\times D_{s_1 t_1}^{\lambda_1 *}(\pi_1)\ D_{s_2 t_2}^{\lambda_2 *}(\pi_2). \tag{45}$$

In writing Eq. (45) we have made use of the direct product property of the $[\lambda_1] \times [\lambda_2]$ IR of $S_{n_1} \times S_{n_2}$ and also made use of the representation property of $[\lambda]$. We now want to perform the sum on $\pi_1 \in S_{n_1}$ and $\pi_2 \in S_{n_2}$. To do so we note that $D^\lambda(\pi_1)$ is irreducible when π_1 is regarded as an element of $S_{n_1+n_2}$ but is reducible when π_1 is regarded as an element of $S_{n_1} \times S_{n_2}$ or just S_{n_1}. In the latter case, $D^\lambda(\pi_1)$ is the direct sum of matrices which are irreducible under S_{n_1}. The Yamanouchi form of the IR matrices readily lends itself to performing the explicit reduction since the matrices are in fact constructed by exploiting the above fact. For example, the $[4, 1]$ IR of S_5 for the element (13) is the direct sum of the IR matrices of S_3, viz.,

$$D^{[41]}(13) = \begin{array}{c} \\ 21111 \\ 12111 \\ 11211 \\ 11121 \end{array} \overset{\displaystyle\begin{array}{cccc} 21111 & 12111 & 11211 & 11121 \end{array}}{\begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & -1/2 & -\sqrt{3}/2 \\ 0 & 0 & -\sqrt{3}/2 & 1/2 \end{pmatrix}}$$

$$= \begin{pmatrix} D^{[3]}(13) & & & 0 \\ & D^{[3]}(13) & & \\ & & D^{[21]}(13) & \\ 0 & & & \end{pmatrix}$$

wherein we explicitly display the reduction. In general then, one may write

$$D_{\alpha\beta}^{\lambda}(\pi_1) = \delta_{\widetilde{\alpha\beta}} \, D_{\underline{\alpha}\underline{\beta}}^{\lambda\overline{\alpha}}(\pi_1). \tag{46}$$

The notation $\delta_{\widetilde{\alpha\beta}}$ of Eq. (46) means that the upper or leftmost n_2 symbols of the Yamanouchi row labels must be the same. That is, the Yamanouchi row label is $r_{m_1+m_2} r_{m_1+m_2-1} \cdots r_2 r_1$, where the subscript, j say, is the number to be placed in the r_j row of a standard Young tableaux. The notation $\underline{\alpha}, \underline{\beta}$ refers to the lower or rightmost n_1 parts of the Yamanouchi symbols which are the row and column labels of the $D^{\lambda\alpha}$ IR of S_{m_1}. The pattern $[\lambda_{\widetilde{\alpha}}]$ is obtained from $[\lambda]$ by removing the α boxes from the $[\lambda]$ pattern. For example, in the [41] IR of S_5 the removal of the 11 from 11211 and 11121 means the removal of two boxes from the [41] pattern to yield the [21] pattern which denotes the [21] IR of S_3, viz.,

The sum over $\pi_1 \in S_{m_1}$ in Eq. (45) may then be carried out by using Eq. (46) and the orthogonality of the IR matrices of S_{m_1}. The result is

$$\sum_{r \in S_{m_1}} D_{sv}^{\lambda}(\pi_1) D_{s_1 t_1}^{\lambda_1 *}(\pi_1)$$

$$= \delta_{\widetilde{sv}} \, \delta_{ss_1} \, \delta_{\underline{v} t_1} \, \delta_{\lambda_1 x_{\widetilde{s}}} n_1! \, / d[\lambda_1] \tag{47}$$

in which the meaning of $\delta_{\lambda_1 x_{\widetilde{s}}}$ is merely that the $[\lambda]$ IR of $S_{m_1+n_2}$ must contain the $[\lambda_1]$ IR of S_{m_1} at least once or the sum vanishes.

At first sight, one might be tempted to use exactly the same technique to perform the π_2 sum of Eq. (45). This would be incorrect, however, since that sum is over permutations on the symbols $n_1 + 1, n_1 + 2 \cdots n_1 + n_2$. The matrices of $S_{m_1+n_2}$ in the Yamanouchi scheme are constructed to be block diagonal for the elements of S_k, $k \leq n_1 + n_2$, with the sequence $S_1, S_2 \cdots S_{m_1+n_2}$. In each case, the permutations involve the first or lowest ordered symbols $1, 2 \cdots k$. Hence, prior to performing the π_2 sum it is necessary to convert the π_2 in $D_{vt}^{\lambda}(\pi_2)$ from a permutation on the symbols $n_1 + 1, n_1 + 2, \ldots,$ $n_1 + n_2$ to a permutation on the first n_2 symbols $1, 2 \cdots n_2$. We denote the $n_1 + 1, n_1 + 2 \cdots n_1 + n_2$ permutation by $\widetilde{\pi}_2$ and the corresponding permutation with $n_1 + j \to j$ by $\underline{\pi}_2$. Since $[\lambda_2]$ is an IR of S_{n_2} it is immaterial whether $\widetilde{\pi}_2$ or $\underline{\pi}_2$ is the argument of $D_{s_2 t_2}^{\lambda_2 *}(\pi_2)$. We require then a permutation x of $S_{m_1+n_2}$ such that

$$\widetilde{\pi}_2 = x^{-1} \underline{\pi}_2 x \tag{48}$$

and such that x is independent of the particular π_2. An adequate choice of x is

$$x = \begin{pmatrix} 1 & 2 \cdots n_1 & n_1 + 1 & n_1 + 2 \cdots n_1 + n_2 \\ n_2 + 1 \cdots n_2 + n_1 & 1 & 2 \cdots n_2 \end{pmatrix}. \tag{49}$$

Only the right part of the lower row, $1, 2 \cdots n_2$, is critical; the left part could be chosen to be any convenient permutation of $n_1 + 1, n_1 + 2, \cdots$. With an appropriate x then

$$D_{vt}^{\lambda}(\widetilde{\pi}_2) = \sum_{\alpha\beta} D_{v\alpha}^{\lambda\widetilde{\alpha}}(x^{-1}) D_{\alpha\beta}^{\lambda}(\underline{\pi}_2) D_{\beta t}^{\lambda}(x). \tag{50}$$

Then

$$\sum_{r_2} D_{vt}^{\lambda}(\pi_2) D_{s_2 t_2}^{\lambda_2}(\pi_2)$$

$$= \sum_{\substack{\alpha\beta \\ r_2}} D_{v\alpha}^{\lambda}(x^{-1}) D_{\alpha\beta}^{\lambda}(\underline{\pi}_2) D_{\beta t}^{\lambda}(x) D_{s_2 t_2}^{\lambda_2 \dagger}(\underline{\pi}_2)$$

$$= \sum_{\alpha\beta} D_{v\alpha}^{\lambda}(x^{-1}) D_{\beta t}^{\lambda}(x) \, \delta_{\overline{\alpha\beta}} \, \delta_{\underline{\alpha} s_2} \, \delta_{\underline{\beta} t_2} \, \delta_{\lambda_{\widetilde{\alpha}} \lambda_2} n_2! \, / d(\lambda_2) \tag{51}$$

in which an overbar denotes the upper n_1 parts of the Yamanouchi label and a undertilde denotes the lower n_2 parts. Also $D_{v\alpha}^{\lambda}(x^{-1}) = D_{\alpha v}^{\lambda *}(x) = D_{\alpha v}^{\lambda}(x)$ because the IR matrices are real unitary. Thus finally Eq. (45) becomes

$$\langle \lambda_1 t_1 \lambda_2 t_2; g_j \, | \, \lambda r : s \lambda_1 s_1 \lambda_2 s_2 \rangle$$

$$= \frac{n_1! \, n_2!}{(n_1 + n_2)!} \frac{d(\lambda)}{d(\lambda_1) d(\lambda_2)} \sum_{\substack{tv \\ \alpha\beta}} D_{rt}^{\lambda}(g_j)$$

$$\times D_{\alpha v}^{\lambda}(x) D_{\beta t}^{\lambda}(x) \, \delta_{\widetilde{sv}} \, \delta_{\underline{v} t_1} \, \delta_{\overline{\alpha\beta}} \, \delta_{\underline{\alpha} s_2} \, \delta_{\underline{\beta} t_2} \tag{52}$$

with the proviso that this transformation bracket vanishes unless $[\lambda]$ contains $[\lambda_1] \times [\lambda_2]$ at least once. To explicitly calculate these brackets we need only know the matrix elements of the elements of S, which consist of products of transpositions and are therefore most easily constructed, and the elements of a single permutation x. If one always takes n_1 to be the larger of n_1, n_2, then instead of the x given in Eq. (49) one could use

$$\begin{pmatrix} 1 & 2 \cdots n_1 - 2 & n_1 - 1 & n_1 & n_1 + 1 \cdots n_1 + n_2 \\ n_1 + 1 \cdots n_1 + n_2 \cdots n_1 - 2 & n_1 - 1 & n_1 & 1 & \cdots & n_2 \end{pmatrix} \tag{53}$$

which is a member of S as well. If $n_1 < n_2$, this choice cannot, of course, be made. With $[\lambda]$ written as $[\lambda_1 \lambda_2 \lambda_3 \cdots \lambda_n]$ and r and s written explicitly as $(r_n r_{n-1} \cdots r_2)$, $(s_n s_{n-1} \cdots s_1)$ respectively, one has for the transposition $(n-1, n)$

$$D_{(r_n \cdots r_1)(s_n \cdots s_1)}^{[\lambda_1 \lambda_2 \cdots \lambda_n]}(n-1, n) = \prod_{j=1}^{n-2} \delta_{r_j s_j}$$

$$\times \begin{cases} \delta_{r_n s_n} \delta_{r_{n-1} s_{n-1}} \delta_{r_n r_{n-1}} \\ \text{and for } r_n \neq r_{n-1} \\ \delta_{r_n s_n} \delta_{r_{n-1} s_{n-1}} \left(\dfrac{1}{\lambda_{r_n} - \lambda_{r_{n-1}} + r_{n-1} - r_n} \right) \\ \delta_{s_n r_{n-1}} \delta_{s_{n-1} r_n} \left(1 - \dfrac{1}{(\lambda_{r_n} - \lambda_{r_{n-1}} + r_{n-1} - r_n)^2} \right)^{1/2} \end{cases} \tag{54a}$$

and for the transposition $(i-1, i)$ in S_n

$$D_{(r_n \cdots r_1)(s_n \cdots s_1)}^{[\lambda_1 \cdots \lambda_n]}(i-1, i)$$

$$= \prod_{j=i+1}^{n} \delta_{r_j s_j} D_{(r_i \cdots r_1)(s_i \cdots s_1)}^{[\lambda_1 \lambda_2 \cdots \lambda_i \, \lambda_{i+1}-1 \, \lambda_{i+2}-1 \cdots \lambda_{n-1}]}(i-1, i) \tag{54b}$$

These readily computerized expressions together with the identity

$$(i, j) = (i, i+1)(i+1, i+2) \cdots (j-1, j)(j-1, j-2) \cdots$$

$$\times (i+1, i) \tag{55}$$

enable one to easily program the transposition brackets.

The overlap of two projected states follows from the

special form of Eq. (38) and is

$$\langle \lambda r' : s' \lambda_1 s_1 \lambda_2 s_2 | \lambda r : s \lambda_1 s_1 \lambda_2 s_2 \rangle$$
$$= \delta_{r'r}[n_1! n_2! / (n_1 + n_2)!][d(\lambda)/d(\lambda_1)d(\lambda_2)]$$
$$\times \sum_\alpha D^\lambda_{\alpha s}(x) D^\lambda_{\alpha s'}(x) \delta_{\underline{s}\,s_1} \delta_{\underline{\alpha}\,s_2}. \tag{56}$$

From the normalization factor which is Eq. (56) with $s' = s$ one has that the projected state vanishes identically unless the lower n_1 entries of s are the same as those of s_1. Hence we could add to Eq. (56) the factor $\delta_{\underline{s}\,'s_1}$ as well.

The condition that two states be identical (when normalized) is

$$\langle \lambda r : s' \lambda_1 s_1 \lambda_2 s_2 | \lambda r : s \lambda_1 s_1 \lambda_2 s_2 \rangle$$
$$= (\langle \lambda r : s \lambda_1 s_1 \lambda_2 s_2 | \lambda r : s \lambda_1 s_1 \lambda_2 s_2 \rangle$$
$$\times \langle \lambda r : s' \lambda_1 s_1 \lambda_2 s_2 | \lambda r : s' \lambda_1 s_1 \lambda_2 s_2 \rangle)^{1/2}. \tag{57}$$

Hence the following is the procedure used to resolve the multiplicity:

(i) Choose the maximal values of s_1 and s_2.

(ii) Compute the matrix whose elements are given by Eq. (56).

(iii) Form the trace of that matrix to determine $n(\lambda, \lambda_1 \times \lambda_2)$.

(iv) Start with maximal s consistent with $\underline{s} = s_1$ labeling the first occurence.

(v) Consider successively the possible s in decreasing order checking the condition of Eq. (57). The first such s' which does not satisfy the condition is to label the second occurrence, etc.

If we return to the $S_3 \supset S_1 \times S_2$ example, we have $[\lambda_1] = [1]$ so that $s_1 = 1$ is the only possibility. Also $[\lambda_2] = [2]$ so $s_2 = 11$. Since $[\lambda] = [21]$, r and s may take on the value 211 and 121. We note that $x = (132)$ and $g_j = e$, (12) or (13). Thus, in abbreviated notation, the transformation bracket of Eq. (52) becomes

$$\langle g_j | r : s \rangle = \tfrac{2}{3} \sum_t D^{[21]}_{rt}(g_j) D^{[21]}_{211s}(132) D^{[21]}_{211r}(132). \tag{58}$$

The overlap matrix is easily found using the matrix representative of (132) which is

$$D^{[21]}(132) = \begin{array}{cc} & \begin{array}{cc} 211 & \quad 121 \end{array} \\ \begin{array}{c} 211 \\ 121 \end{array} & \begin{pmatrix} -1/2 & \sqrt{3}/2 \\ \sqrt{3}/2 & -1/2 \end{pmatrix}, \end{array}$$

which follows from $(132) = (23)(12)$ and the IR matrices of (12) and (23) which follow from Eqs. (54) as

$$D^{[21]}(12) = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}, \quad D^{[21]}(23) = \begin{pmatrix} -1/2 & \sqrt{3}/2 \\ \sqrt{3}/2 & 1/2 \end{pmatrix}.$$

Thus the overlap matrix given by Eq. (56) becomes

$$\langle s' | s \rangle = \tfrac{2}{3} D^{[21]}_{211s}(132) D^{[21]}_{211s'}(132)$$
$$= \tfrac{1}{6}\begin{pmatrix} 1 & \sqrt{3} \\ \sqrt{3} & 3 \end{pmatrix}.$$

One sees immediately that the state projected using $s' = 121$ is the same as that using $s = 211$. Indeed, if one

normalizes the projected state, the normalized transformation bracket $\langle g_j | r : s \rangle / \sqrt{\langle s | s \rangle}$ is given by

$$\langle g_j | r : s \rangle / \sqrt{\langle s | s \rangle} = \sqrt{2/3} \sum_t D^{[21]}_{rt}(g_j) D^{[21]}_{211t}(132),$$

which is independent of s. If one successively inserts $g_j = e$, (12) and (13) for $r = 211$, one finds that the projected state differs from the "brute force" state of Eq. (12) only by a trivial overall minus sign.

IV. THE SIMPLEST EXAMPLE WITH MULTIPLICITY 2

The simplest example in which an IR of $S_{n_1+n_2}$ occurs more than once in a representation induced from an IR of some decomposition of $n_1 + n_2$ is the case $S_6 \supset S_3 \times S_3$, where the $[21] \times [21]$ IR of $S_3 \times S_3$ induces the $[321]$ IR of S_6 twice. In this case, the possible choices of s_1 and s_2 are 211 or 121. The dimension of $[321]$ is 16, but the condition that $\underline{s} = s_1$ leaves but 12 of these as valid potential multiplicity labels for each choice of s_1. If, instead of fixing s_1 and s_2 by the choice indicated earlier, one considers all possible choices together with all possible consistent values of s, the overlap matrix has dimension 24. When the normalization factors are inserted, one finds that there are but four distinct (not identical) state labelings. That is, for example, the states projected from $s_1 = 211$, $s_2 = 211$ with $s = 321\ 211$ are identical with the states projected from $s_1 = 211$, $s_2 = 211$ with $s = 312\ 211$. But these states are distinct (their overlap is not 1) from

s_1	s_2	s
211	211	231211
211	211	132211
211	121	132211.

In addition, these above states are distinct from one another. Hence with this scheme, there are but four distinct state labelings of which only two can be linearly independent since the multiplicity is but 2. An alternate resolution of the multiplicity is to diagonalize this 4×4 overlap matrix which, of course, has two zero eigenvalues. The corresponding eigenvectors give then the weighting coefficients for each of the four distinct state labelings. But this is not without ambiguity on account of the multiplicity of the zero eigenvalues, which means that the corresponding eigenvectors are not uniquely determined.

Within the procedure we have outlined, the first occurrence of the $[321]$ IR is projected using $s = 321211$ from the state labelled by $s_1 = 211$, $s_2 = 211$. The second occurrence is projected using $s = 231211$. The projected states from the two occurrences are not orthogonal this way, but they are uniquely determined, and their overlap is minimized insofar as the use of any other pair of the four distinct states is concerned.

It is to be noted that we have not proved that our scheme always gives minimal overlap of the multiple occurrences. Indeed we must confess our inability to do so in the general case. Nonetheless, the use of induced representation techniques together with our resolution of the multiplicity provides a complete labeling scheme and a "canonical" resolution of the multiplicity. These

states may now be used to form appropriate coefficients of fractional parentage and useful calculations performed with them. In any physical problem, of course, this arbitrariness in the multiplicity resolution does not show up in the final answer. The scheme presented here is admittedly aimed at computerization of the process, and it is with regret that we note that a truly elegant resolution of the multiplicity remains an unsolved problem.

[1]E. P. Wigner, Phys. Rev. 51, 106 (1937).
[2]H. A. Jahn and H. Van Wieringen, Proc. Roy. Soc. A 209, 502 (1951).
[3]J. P. Elliott, J. Hope, and H. A. Jahn, Proc. Roy. Soc. A 246, 241 (1953).
[4]P. Kramer and M. Moshinsky, "Group Theory of Harmonic Oscillators and Nuclear Structure," in *Group Theory and Its Applications*, edited by Ernest M. Loebl (Academic, New York, 1968).
[5]J. P. Elliott, Proc. Roy. Soc. A 245, 128, 562 (1958); J. P. Elliott and M. Harvey, Proc. Roy. Soc. A 272, 557 (1962).
[6]J. P. Draayer, D. L. Pursey, and S. A. Williams, Nucl. Phys. A 119, 577 (1968); J. P. Draayer and S. A. Williams, Nucl. Phys. A 129, 647 (1969).
[7]J. P. Draayer, J. Math. Phys. 11, 3225 (1970).
[8]Morton Hammermesh, *Group Theory and Its Application to Physical Problems* (Addison-Wesley, Reading, Mass., 1962).
[9]U. Fano and G. Racah, *Irreducible Tensorial Sets* (Academic, New York, 1959).
[10]F. D. Murnaghan, *The Theory of Group Representation* (John Hopkins Press, Baltimore, 1938).
[11]We use the standard Young labeling of the IR of S_n and the Yamanouchi labeling of the rows as given by Ref. 8.
[12]J. S. Lomont, *Applications of Finite Groups* (Academic, New York, 1959).

Symmetry and invariance properties of the Boltzmann equation on different groups

J. N. Massot* and R. Bacis†

Laboratoire de Physique du Rayonnement, Université de Saint-Etienne, 23 Rue du Docteur Paul Michelon, 42100-Saint-Etienne, France
(Received 6 October 1975)

The introduction of the group theory in the treatment of the Boltzmann equation shows the reducibility of the collision integral operator on the invariant subspaces of Klein V or SO_2 group. Especially we prove the equality of matrices representing the collision integral operator between inequivalent subspaces first in its linear form and then in its general form. These results are finally expanded to the full Boltzmann equation when we consider its properties as a whole in the phase space $(\mathcal{E}_r \times \mathcal{E}_v)$. This brings back Boltzmann equation following the Chapmann–Enskog process to the differential equation system depending solely on the variable $|r|$. The examination of the Boltzmann equation symmetries allows us to obtain the selection rules which lead to an important simplification in theoretical as well as numerical calculations of the distribution function.

1. INTRODUCTION

It is well known that group theory makes an important contribution both towards our understanding of a certain number of physical processes and towards research conducted on their solutions. This is the case of atomic or molecular spectroscopy and nuclear physics. The majority of positive results obtained are linked with the demonstration of the symmetries of the basic equations of these processes (Schrödinger, Dirac, Klein–Gordon, . . ., equations).

We propose to apply the same technique to the Boltzmann equation. The structure of this equation is well known and is the subject of many studies. A considerable bibliography is to be found in the works of Chapman and Cowling,[1] Hirsfelder *et al.*[2] and, more recently, Ferziger and Kaper.[3]

However, in view of our aim, we feel it is particularly important to call attention to the work of Kumar.[4-6] This author exploits, by algebraic methods, the invariance property under rotation of the Boltzmann collision operator. This enables him to express this operator in a spherical coordinate basis in the space $\mathcal{E}_v$ of the velocities which is in fact a standard basis for rotation group representations. Consequently, it is possible for him to simplify the description of the Chapman–Enskog process, used for research on the solutions of the Boltzmann equation, and to show that this equation can be decomposed into irreducible tensorial operators in relation to the rotation group of $\mathcal{E}_v$.

The expansion obtained by Kumar is thus linked both to the invariance property under rotation of the collision operator and to the choice of a determined basis, the standard basis of the rotation group. However, other bases may be used to describe the distribution function $f(r, v, t)$, solution to the Boltzmann equation, and particularly bases of cylindrical or Cartesian symmetry. Indeed the choice of the basis is especially influenced by the boundary conditions of the problem when these conditions are taken into account and by the symmetry of forces F applied to the particles.

The collision operator has a well determined reduc-

ible structure in each of these bases. This structure corresponds to an invariance of the operator with respect to a particular group of transformations and gives rise to precise selection rules which is essential to know for the calculations.

In order to demonstrate these results and to show the nature of the transformation groups under which the collision operator is invariant, we shall use a more rapid method than that employed by Kumar in the case of a basis with spherical coordinates. In this work, we shall directly apply theorems arising from the group theory and in particular the Wigner–Eckart theorem. The demonstration will be made when the operator is linear, then when it is expressed in its general form. Owing to its reducibility the operator is expressed according to the direct sum of matrices associated with the irreducible representations of the invariance groups which are basically inequivalent. We shall show in fact that certain of them are necessarily equal and we shall complete this study by indicating the symmetry properties of the matrix elements. This work is the subject of Secs. 2 and 3, for the invariants with respect to the Klein V group, and of Sec. 4 for those concerning the SO_2 group.

All the above properties are those which result only from the study of the expression of the collision operator in the velocity space $\mathcal{E}_v$. One can go further and consider the phase space $(\mathcal{E}_r \times \mathcal{E}_v)$ as a whole. In these circumstances, it is possible to show that the Boltzmann equation possesses invariance properties which are determined according to the symmetries of the F force applied to the particles of the system. In order to exploit these properties, it is necessary to refer to the results expressed in Secs. 2, 3, and 4. Especially, besides demonstrating the selection rules and invariant subspaces when the F force is invariant by rotation or zero, we express the Boltzmann equation eigenfunctions in the subspace $(\mathcal{E}_r^* \times \mathcal{E}_v^*)$, which is the tensorial product of the spaces corresponding to the angular parts of the r and v variables. This brings the Boltzmann equation back to a differential equation system corresponding to the variable $|r|$ to which the Chapman–Enskog process is still applicable. Section 5 will be devoted to

these questions and it will be shown that the moments of the distribution function expressed in a local basis are finite linear combinations of the moments connecting to an absolute basis.

2. REDUCIBILITY OF THE INTEGRAL COLLISION OPERATOR IN THE HERMITE POLYNOMIAL BASIS

A. Definitions and elementary properties of the linearized integral

The calculation of the collision integral in its linearized form was achieved by Kumar.[4-6] By following the traditional notations[7] and limiting ourselves to the case of a single gas in order to simplify the expression, we have

$$D[\phi_1] = \int f_1^0 f_2^0 V \sigma(v, \Omega') [\Delta \phi_{12}] d\Omega' \, d\mathbf{v}_2 \qquad (2.1)$$

where Ω' represents the polar angles of the vector $\mathbf{V}' = \mathbf{v}'_1 - \mathbf{v}'_2$, relative velocity vector of the particles 1 and 2 after the collision, and where V is the modulus of this same vector before the collision. Finally $\sigma(v, \Omega')$ is the effective differential cross section of the colliding particles and

$$\Delta \phi_{12} = \phi'_1 + \phi'_2 - \phi_1 - \phi_2, \qquad (2.2)$$

the functions ϕ_i being deduced from the distribution function f_i by the well-known relation

$$f_i = f_i^0 (1 + \phi_i), \qquad (2.3)$$

and f_i^0 being the Maxwell distribution.

Under these circumstances, it is possible to evaluate the collision operator D in spherical basis $Q_{nlm} = R_{nl} Y_{lm}$ (see definition in Appendix A) which gives

$$\int Q_{n'l'm'}^* D[Q_{nlm}] d\mathbf{v}_1 = \langle n'l'm' | D | nlm \rangle. \qquad (2.4)$$

In the absence of any polarization of the physical system constituted by the particles 1 and 2, the effective differential cross section σ depends solely on the relative angle between the vectors $\mathbf{V}'$ and $\mathbf{V}$, and the operator D is invariant under rotation. It is then simple to see that the matrix elements (2.4) are diagonal in l and m by using the Wigner—Eckart theorem[8,9] which gives, in this case,

$$\langle n'l'm' | D | nlm \rangle = \delta_{ll'} \delta_{mm'} D_{n'n}^l. \qquad (2.5)$$

The explicit calculation of the matrix elements $D_{n'n}^l$ was effected in detail by Kumar[4,5] and the corresponding result is found in the relations $(103)^4$ and $(121).^5$

B. Reducibility of the collision operator in the Hermite polynomial basis

In this section our aim is to show that the operator D is still reducible in the basis $|n_x n_y n_z\rangle$ spanned by the Hermite polynomials $H_\mathbf{n}(\mathbf{v}) = H_{n_x}(v_x) H_{n_y}(v_y) H_{n_z}(v_z)$ $(n_x, n_y, n_z$ integers $\geqslant 0)$. These polynomials are defined in Appendix A. They differ from those used by Grad[10] in that they are immediately factorizable according to the variables (v_x, v_y, v_z) which simplifies their use.

The two bases $|nlm\rangle$ and $|n_x n_y n_z\rangle$ are bases of the L^2 Hilbert space of square-integrable functions. Thus there exists a unitary transformation with coefficients $\langle nlm | n_x n_y n_z \rangle$ such that

$$H_\mathbf{n} = \sum_{nlm} \langle nlm | n_x n_y n_z \rangle Q_{nlm} \qquad (2.6)$$

with

$$\sum_\mathbf{n} \langle n'l'm' | n_x n_y n_z \rangle \langle n_x n_y n_z | nlm \rangle = \delta_{nn'} \delta_{ll'} \delta_{mm'} \qquad (2.7)$$

$$\sum_{nlm} \langle n'_x n'_y n'_z | nlm \rangle \langle nlm | n_x n_y n_z \rangle = \delta_{n_x n'_x} \delta_{n_y n'_y} \delta_{n_z n'_z}. \qquad (2.8)$$

This transformation has been studied by many authors and especially by Domergue.[11] The explicit expression of $\langle nlm | \mathbf{n} \rangle$ can be found in Appendix B.

Then, we can evaluate the matrix elements of D in the Cartesian basis,

$$\int H_{\mathbf{n}'} D[H_\mathbf{n}] d\mathbf{v}_1 = \langle \mathbf{n}' | D | \mathbf{n} \rangle = D_{\mathbf{n}'\mathbf{n}}, \qquad (2.9)$$

where by using (2.7), (2.8), and (2.5),

$$D_{\mathbf{n}'\mathbf{n}} = \sum_{nn'lm} \langle \mathbf{n}' | n'lm \rangle D_{n'n}^l \langle nlm | \mathbf{n} \rangle. \qquad (2.10)$$

With selection rules [Appendix B, Eq. (B2)] we can see that

$$(-)^{n_x + n_y + n_y} = (-)^{n'_x + n'_y + n'_z} \qquad (2.11)$$

and

$$(-)^{n'_z} = (-)^{n_z}. \qquad (2.12)$$

On the other hand, the matrix elements $D_{n'n}^l$ and $D_{\mathbf{n}'\mathbf{n}}$ being real and the coefficients $\langle n_x n_y n_z | nlm \rangle$ being real or imaginary according to the parity of n_y, it is required that $(n_y + n'_y)$ be even. Therefore,

$$(-)^{n_y} = (-)^{n'_y}. \qquad (2.13)$$

Finally, with S the signature $\{(-)^{n_x}, (-)^{n_y}, (-)^{n_z}\}$ we have

$$S = S'. \qquad (2.14)$$

Thus

$$D_{\mathbf{n}\mathbf{n}} = D_{(\mathbf{n}),(\mathbf{n}')}^S \delta_{SS'}, \qquad (2.15)$$

where the notation $(\mathbf{n})$ means that the indices $(n_x n_y n_z)$ are taken in such a way that S be constant.

Thus we have proved that D was reducible in the $\{H_\mathbf{n}\}$ basis. The number of possible signatures S being finite and equal to eight, the number of invariant subspaces and therefore the number of submatrices, which result in the reduction of D, is finite and equal to eight. The complete set of S values is

$$\{S\} = \{(+++), (++-), (+-+), (-++), (+--),$$
$$(-+-), (--+), (---)\}. \qquad (2.16)$$

C. Dimension of the invariant subspaces for a basis of finite dimension

The reducibility examination would not be complete if we did not compute the dimensions of each submatrix when the basis is of finite dimension, equal to the numbers of independent polynomials $H_\mathbf{n}$ of maximum degree N_M. We shall show that one relation is enough to express the dimensions connecting to the eight possible signatures.

The submatrices D^S of elements $D_{(\mathbf{n})(\mathbf{n}')}^S$ have dimensions which can be deduced from all the possible values

of the indices n_x, n_y, $n_z \in [0, N_M]$ for the signature S and such that

$$n_x + n_y + n_z \leqslant N_M. \tag{2.17}$$

Let us consider first of all the signature $(+ + +)$.

The indices n_x, n_y, n_z are necessarily even and equal to $2k_x$, $2k_y$, $2k_z$, so that

$$2k_x + 2k_y + 2k_z \leqslant 2K, \tag{2.18}$$

where $K = [N_M/2]$ is the integer part of $N_M/2$.

Thus the dimension of the submatrix $(+ + +)$ is given by

$$\dim[D^{(+++)}] = d(K) = \sum_{k_x + k_y + k_z \leqslant K} 1 = \tfrac{1}{6}[K^3 + 6K^2 + 11K + 6]. \tag{2.19}$$

After some calculations, the other signatures give the same relation $d(K)$ with

$$K = \left[\frac{N_M - k}{2}\right], \tag{2.20}$$

where

$$k = \begin{cases} 1 & \text{if } S \in \{(+ + -), (+ - +), (- + +)\}, \\ 2 & \text{if } S \in \{(+ - -), (- + -), (- - +)\}, \\ 3 & \text{if } S = (- - -). \end{cases} \tag{2.21}$$

As an example, we give in Fig. 1 the structure of D with a basis of 20 polynomials $(N_M = 3)$.

3. STUDY OF D IN THE CARTESIAN BASIS $H_{\vec{n}}$

A. Groups leading to the reducibility of D

The above calculations are of an algebraic type. They
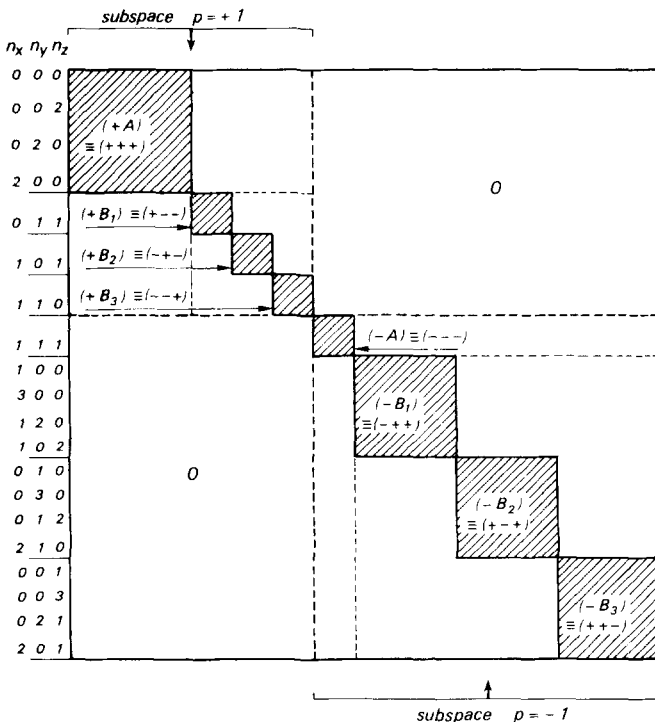
FIG. 1. Representation of the integral Boltzmann operator in the Cartesian basis $H_{\vec{n}}^{p\Gamma}$ for $N_M = 3$.

enable us to demonstrate the reducibility of D in the $\{H_{\vec{n}}\}$ basis, but their possibilities are limited. In particular, they cannot give us the physical reasons for the existence of this reducibility.

Thus, the aim of this section is to answer this question. In so doing, we shall obtain new results such as those concerning the equality of the matrices $D^{(++-)}$, $D^{(+-+)}$, and $D^{(-++)}$ on one hand, and $D^{(--+)}$, $D^{(-+-)}$, $D^{(+--)}$ on the other hand. Moreover, we shall enlarge these results to the general case of the operator in its nonlinear form. For this, we shall show that D commutes with all the operators of the Klein finite V group.

Let us remember first of all that D commutes with the operators R of the rotation group and also with the inversion operator of space P. So let us consider the operators I_x, I_y, and I_z defined as the inversion operators of axis x, y, z. These operators commute

$$[I_x, I_y] = [I_y, I_z] = [I_z, I_x] = 0, \tag{3.1}$$

and we have

$$\begin{aligned} I_x H_{\vec{n}} &= I_x H_{n_x}(v_x) H_{n_y}(v_y) H_{n_z}(v_z) \\ &= H_{n_x}(- v_x) H_{n_y}(v_y) H_{n_z}(v_z). \end{aligned} \tag{3.2}$$

But the Hermite polynomials have a defined parity. Thus,

$$H_{n_x}(- v_x) = (-)^{n_x} H_{n_x}(v_x). \tag{3.3}$$

So, in the general case

$$I_i | n_x n_y n_z \rangle = (-)^{n_i} | n_x n_y n_z \rangle \quad \forall \, i \in [x, y, z]. \tag{3.4}$$

Thus the states $| n_x n_y n_z \rangle$ are eigenstates of the operators I_i. Also, they are eigenstates of the operator P, for

$$P | n_x n_y n_z \rangle = I_x I_y I_z | n_x n_y n_z \rangle = (-)^{n_x + n_y + n_z} | n_x n_y n_z \rangle. \tag{3.5}$$

This operator is peculiar to an Abelian group, isomorphic to an S_2 group possessing two irreducible representations,

$$p = \pm 1 \quad \text{with } p = (-)^{n_x + n_y + n_z}.$$

Let us consider then the three operators R_i, such that

$$R_i = I_j I_k, \quad (i, j, k) \text{ permutation from } (x, y, z). \tag{3.6}$$

If we denote by E the operator identity, we can see that these three operators form with E a finite group which is easily identified[12,13] with the Klein V group and we have

$$R_i R_j = R_k, \quad E R_i = R_i, \quad R_i R_i = E. \tag{3.7}$$

Moreover, these operators commute,

$$[R_i, R_j] = 0. \tag{3.8}$$

We know that this group is Abelian and that there are four irreducible representations of dimension one (A, B_1, B_2, B_3), the characters of which are given by

Γ	E	R_x	R_y	R_z
A	1	1	1	1
B_1	1	1	-1	-1
B_2	1	-1	1	-1
B_3	1	-1	-1	1

The action of the operators R_i is the same as a directional change of the axis (j,k). It is therefore equivalent to a rotation of the axis i by the angle $+\pi$. Then these operators are rotation operators which commute with the collision operator D. Thus

$$[D, R_i] = 0 \qquad i \in [x, y, z]. \tag{3.9}$$

On the other hand, we have

$$[R_i, P] = 0, \tag{3.10}$$

for P and R_i are made with I_i which commute between them. If we study the action of the operators R_i on the basis $|n_x n_y n_z\rangle$, we have, bearing (3.4) and (3.6) in mind,

$$R_i \,|\, n_x n_y n_z\rangle = I_j I_k \,|\, n_x n_y n_z\rangle = (-)^{n_j + n_k} \,|\, n_x n_y n_z\rangle. \tag{3.11}$$

Thus the basis $|n_x n_y n_z\rangle$ is the basis belonging to the operators R_i. It therefore defined a standard basis for the irreducible representations of R_i and the same holds true for the operator P, given the relation (3.5). As the operators (P, R_x, R_y, R_z) commute between each other, it is possible to classify the states $|n_x n_y n_z\rangle$ according to the irreducible representations of P and R_i.

If we classify the states with the help of the eigenvalue p of P and the representation $\Gamma \in \{A, B_1, B_2, B_3\}$, denoting $|n_x n_y n_z\rangle$ by $|(\mathbf{n}), p, \Gamma\rangle$, we have

$$P \,|\, (\mathbf{n}), p, \Gamma\rangle = p \,|\, (\mathbf{n}), p, \Gamma\rangle = (-)^{n_i + n_j + n_k} \,|\, (\mathbf{n}), p, \Gamma\rangle, \tag{3.12}$$

$$R_i \,|\, (\mathbf{n}), p, \Gamma\rangle = (-)^{n_j + n_k} \,|\, (\mathbf{n}), p, \Gamma\rangle. \tag{3.13}$$

The examination of the characters of the R_i representations in the basis $|(\mathbf{n}), p, \Gamma\rangle$ enables us to identify the irreducible representations. Thus we obtain the mapping

Γ	$S(p = +1)$	$S(p = -1)$
A	$(+ + +)$	$(- - -)$
B_1	$(+ - -)$	$(- + +)$
B_2	$(- + -)$	$(+ - +)$
B_3	$(- - +)$	$(+ + -)$

$$\tag{3.14}$$

We can therefore connect two distinct values of the signature S with each representation, every one of them belonging to the two eigenvalues ± 1 of P.

Thus, we can write $|(\mathbf{n}), p, \Gamma\rangle = |(\mathbf{n}), S\rangle$. At last, we have

$$\langle(\mathbf{n}), p, \Gamma \,|\, P \,|\, (\mathbf{n}'), p', \Gamma'\rangle = \delta_{(\mathbf{n})(\mathbf{n}')} \delta_{\Gamma\Gamma'} \delta_{pp'} P^p,$$

$$\langle(\mathbf{n}), p, \Gamma \,|\, R_i \,|\, (\mathbf{n}'), p', \Gamma'\rangle = \delta_{(\mathbf{n})(\mathbf{n}')} \delta_{pp'} \delta_{\Gamma\Gamma'} R_i^\Gamma. \tag{3.15}$$

Then, according to the Wigner—Eckart theorem, we have

$$\langle(\mathbf{n}) p \Gamma \,|\, D \,|\, (\mathbf{n}') p' \Gamma'\rangle = \delta_{pp'} \delta_{\Gamma\Gamma'} D_{(\mathbf{n})(\mathbf{n}')}^{p\Gamma}. \tag{3.16}$$

The number of possible values for couple (p, Γ) is equal to the product of the numbers of irreducible representations of the S_2 and V groups, i.e., $(2 \times 4) = 8$. Thus, the D operator is broken into a direct sum of eight matrices belonging to the eight invariant subspaces which then can be labelled by (S) or (p, Γ).

B. Equality of the matrices of the subspaces $\Gamma = B_i$

A priori the eight matrices $D^{p\Gamma} = D^S$ are neither iden-

tical nor equivalent, because the irreducible representations are inequivalent.

However, the results obtained in Sec. 2, Part C show that the dimension of (p, B_i) invariant subspaces is independent of i value, in each space of finite dimension N_M. Thus, we may ask whether the resulting submatrices D^{pB_i} are equivalent. Actually, we shall see that they are identical.

For this, the commutation relation $[D, R] = 0$ gives

$$\langle(\mathbf{n}) p \Gamma \,|\, D \,|\, (\mathbf{n}') p \Gamma\rangle = \langle(\mathbf{n}) p \Gamma \,|\, R^{-1} D R \,|\, (\mathbf{n}') p \Gamma\rangle. \tag{3.17}$$

Generally, the action of any operator R on a state $|(\mathbf{n}), p \Gamma\rangle$ does not give rise to a state (or a linear combination of states) of the same subspace (p, Γ) for this is not invariant with regard to group rotation operations. Thus, the relation (3.17) expresses an equality of matrix elements belonging to different subspaces. However, the expression (3.17) is too general to be easily exploitable. Let us consider therefore the particular rotation operators $R_0^\pm$ by $\pm 2\pi/3$ about the ternary symmetry principal axis passing through the origin O of the frame $(Oxyz)$.

So it is obvious that this operation permutes the axis in a direct or inverse cyclic way. Thus

$$R_0^+ \,|\, n_x n_y n_z\rangle = \,|\, n_z n_x n_y\rangle \tag{3.18}$$

and finally, as this operation conserves the parity, we find

$$R_0^+ \,|\, (n_x n_y n_z) p, A\rangle = \,|\, (n_z n_x n_y) p, A\rangle,$$

$$R_0^+ \,|\, (n_x n_y n_z) p, B_i\rangle = \,|\, (n_z n_x n_y) p, B_j\rangle, \tag{3.19}$$

(i, j, k) cyclic permutation from $(1, 2, 3)$.

Thus we establish that the subspace $\Gamma = A$ is invariant with regard to the action of R_0. On the other hand the subspaces (B_1, B_2, B_3) are permuted circularly. By using these results with (3.18), we have

$$\langle(n_x n_y n_z) p A \,|\, D \,|\, (n_x' n_y' n_z') p A\rangle$$

$$= \langle(n_z n_x n_y) p A \,|\, D \,|\, (n_z' n_x' n_y') p A\rangle \tag{3.20}$$

$$= \langle(n_y n_z n_x) p A \,|\, D \,|\, (n_y' n_z' n_x') p A\rangle,$$

$$\langle(n_x n_y n_z) p B_1 \,|\, D \,|\, (n_x' n_y' n_z') p B_1\rangle$$

$$= \langle(n_z n_x n_y) p B_2 \,|\, D \,|\, (n_z' n_x' n_y') p B_2\rangle$$

$$= \langle(n_y n_z n_x) p B_3 \,|\, D \,|\, (n_y' n_z' n_x') p B_3\rangle. \tag{3.21}$$

Thus the matrix elements D^{pA} are dependent since the two relations (3.20) exist between them.

On the other hand, the matrix elements of the D^{pB_1}, D^{pB_2}, and D^{pB_3} matrices are identical according to (3.21). Finally, we can say that in the Hermite polynomial basis, D is the direct sum of the eight matrices $D^{p\Gamma}$ with

$$D^{pB_1} = D^{pB_2} = D^{pB_3}, \quad p = \pm 1. \tag{3.22}$$

C. General case of the nonlinear collision operator

In this case, the operator D is such that

$$D[f_1 f_2] = \int V\sigma(v, \Omega')[f_1' f_2' - f_1 f_2] d\Omega' \, dv_2. \tag{3.23}$$

Its matrix elements have been given on the basis

$|nlm\rangle$ by Kumar[5] (relations 85,91). Formally, these are written

$$\int Q^*_{nlm} D[Q_{n_1 l_1 m_1} Q_{n_2 l_2 m_2}] d\mathbf{v}_1$$

$$= \langle nlm | D | n_1 l_1 m_1, n_2 l_2 m_2 \rangle. \tag{3.24}$$

In this form, the matrix D is not reduced according to the invariant subspaces under the rotation group, since the basis $|n_1 l_1 m_1, n_2 l_2 m_2\rangle$ is not a standard basis. Nevertheless, it is a complete basis for the operations of the group and it is possible to construct a standard basis on it by introducing the Clebsch—Gordan co-efficients[8] $\langle l_1 m_1 l_2 m_2 | lm \rangle$ of the group.

Thus, D must be expanded on the standard basis $Q_{lm}(n_1 n_2 l_1 l_2)$ such that

$$Q_{lm}(n_1 n_2 l_1 l_2) = \sum_{m_1 m_2} \langle l_1 m_1 l_2 m_2 | lm \rangle Q_{n_1 l_1 m_1} Q_{n_2 l_2 m_2} \tag{3.25}$$

whence, in this basis,

$$\int Q^*_{lm}(n0l0) D[Q_{l'm'}(n_1 n_2 l_1 l_2)] d\mathbf{v}_1$$

$$= \langle (n0l0)lm | D | (n_1 n_2 l_1 l_2)l'm' \rangle. \tag{3.26}$$

In this form, the Wigner—Eckart theorem can be applied and we have

$$\langle (n0l0)lm | D | (n_1 n_2 l_1 l_2)l'm' \rangle = \delta_{ll'} \delta_{mm'} D^l_{(n0l0),(n_1 n_2 l_1 l_2)}. \tag{3.27}$$

We note that this result does not explicitly appear in Kumar's work because he uses the nonstandard basis $\{Q_{n_1 l_1 m_1}, Q_{n_2 l_2 m_2}\}$ and this is why his matrix elements directly depend on the indices (m_1, m_2, m) by the use of Clebsch—Gordan coefficients.

The above technique can be used when the basis considered is the $\{H_n\}$ basis. However, it requires the construction of the Clebsch—Gordan coefficients for the irreducible representations (p, Γ). We will note these by $\langle p_1 \Gamma_1, p_2 \Gamma_2 | p_3 \Gamma_3 \rangle$. The irreducible representations of P or $\{R_i\}$ being of dimension one, the Clebsch—Gordan coefficients are equal to 0 or to 1 following the selection rules described below. These selection rules are obtained from the characters of the expansion in irreducible parts of the tensorial product $D^{p_1 \Gamma_1} \times D^{p_2 \Gamma_2}$. We necessarily find $p_3 = p_1 p_2$. The value of Γ_3 is given in the following table:

Γ_1 \ Γ_2	A	B_1	B_2	B_3
A	A	B_1	B_2	B_3
B_1	B_1	A	B_3	B_2
B_2	B_2	B_3	A	B_1
B_3	B_3	B_2	B_1	A

$$\tag{3.28}$$

Then, we can introduce the standard basis $H^{p\Gamma}((n_1)p_1 \Gamma_1, (n_2)p_2 \Gamma_2)$ of the group by the relation

$$H^{p\Gamma}((n_1)p_1 \Gamma_1, (n_2)p_2 \Gamma_2) = \langle p_1 \Gamma_1, p_2 \Gamma_2 | p\Gamma \rangle H^{p_1 \Gamma_1}_{(n_1)} H^{p_2 \Gamma_2}_{(n_2)}. \tag{3.29}$$

Here we do not have the summation over indices such as m_1 and m_2 as in the relation (3.25) because the irreducible representations of S_2 or of the Klein V group

have dimension one. The matrix elements of D in the standard basis are so given by

$$\int H^{p\Gamma}(\alpha) D H^{p'\Gamma'}(\alpha') d\mathbf{v}_2 = \langle \alpha, p\Gamma | D | \alpha', p'\Gamma' \rangle = \delta_{pp'} \delta_{\Gamma\Gamma'} D^{p\Gamma}_{\alpha\alpha'}, \tag{3.30}$$

where we applied the Wigner—Eckart theorem and we denoted by α and α' the indice sets $[(n)p\Gamma, (0)1A]$ and $[(n_1)p_1 \Gamma_1, (n_2)p_2 \Gamma_2]$.

The relation (3.30) therefore shows the existence of invariant subspaces for the collision operator when we consider it in its general nonlinear form and these subspaces are obviously identical and equal in number to those obtained when D is linearized. Moreover, the submatrices D^{pB}_i with the same p value are equal to each other according to the result (3.21). Likewise (3.20) has to be used for the calculation of matrix elements of D^{pA}.

D. Reduction in irreducible tensors

To conclude, we must note that, because of the existence of eight invariant subspaces, the distribution function can be decomposed in a sum of eight functions in connection with each subspace,

$$f(\mathbf{r}, \mathbf{v}, t) = \sum_{p, \Gamma} f^{p\Gamma}(\mathbf{r}, \mathbf{v}, t). \tag{3.31}$$

Especially, on the $\{H_n\}$ basis we have

$$f^{p\Gamma}(\mathbf{r}, \mathbf{v}, t) = \sum_{(n)} F^{p\Gamma}_{(n)}(\mathbf{r}, t) H^{p\Gamma}_{(n)}(\mathbf{v}) \tag{3.32}$$

with

$$F^{p\Gamma}_{(n)}(\mathbf{r}, t) = \int W H^{p\Gamma}_{(n)} f^{p\Gamma} d\mathbf{v} = \int W H^{p\Gamma}_{(n)} f d\mathbf{v}, \tag{3.33}$$

where we usually have[7]

$$W(v) = \frac{f_0(\mathbf{r}, \mathbf{v}, t)}{n(\mathbf{r}, t)} = \left(\frac{\beta m}{2\pi}\right)^{3/2} \exp(-\beta m v^2/2). \tag{3.34}$$

The relations (3.31)—(3.33) enable us to expand any function (and any distribution moment) in irreducible functions of the $(S_2 \times V)$ group leading to collision operator symmetries.

On the other hand, evaluation of irreducible tensorial operators can be done using the Wigner—Eckart theorem. If $T^{p\Gamma}$ is an irreducible operator of the $(S_2 \times V)$ group, this theorem still gives

$$\int W H^{p_1 \Gamma_1}_{(n_1)} T^{p\Gamma} H^{p_2 \Gamma_2}_{(n_2)} d\mathbf{v}$$

$$= \langle (n_1) p_1 \Gamma_1 \| T^{p\Gamma} \| (n_2) p_2 \Gamma_2 \rangle \langle p\Gamma, p_2 \Gamma_2 | p_1 \Gamma_1 \rangle^*. \tag{3.35}$$

The Clebsch—Gordan $\langle p\Gamma, p_2 \Gamma_2 | p_1 \Gamma_1 \rangle$ is real and equal to 0 or 1. The reduced matrix element is therefore equal to the integral if the selection rule imposed by the Clebsch—Gordan coefficient is satisfied, i.e., if $p_1 = p p_2$ and if the representations Γ, Γ_2, and Γ_1 conform to table (3.28).

The results corresponding to these selection rules are given in Fig. 2 for a subspace of maximum finite dimension $N_M = 3$. In each subspace, we have given the (p, Γ) value of the irreducible tensors which are not

$n_x n_y n_z$	+A	+B₁	+B₂	+B₃	-A	-B₁	-B₂	-B₃
+A { 000 / 002 / 020 / 200 }	(+A)	(+B₁)	(+B₂)	(+B₃)	(-A)	(-B₁)	(-B₂)	(-B₃)
+B₁ 011	(+B₁)	(+A)	(+B₃)	(+B₂)	(-B₁)	(-A)	(-B₃)	(-B₂)
+B₂ 101	(+B₂)	(+B₃)	(+A)	(+B₁)	(-B₂)	(-B₃)	(-A)	(-B₁)
+B₃ 110	(+B₃)	(+B₂)	(+B₁)	(+A)	(-B₃)	(-B₂)	(-B₁)	(-A)
-A 111	(-A)	(-B₁)	(-B₂)	(-B₃)	(+A)	(+B₁)	(+B₂)	(+B₃)
-B₁ { 100 / 300 / 120 / 102 }	(-B₁)	(-A)	(-B₃)	(-B₂)	(+B₁)	(+A)	(+B₃)	(+B₂)
-B₂ { 010 / 030 / 012 / 210 }	(-B₂)	(-B₃)	(-A)	(-B₁)	(+B₂)	(+B₃)	(+A)	(+B₁)
-B₃ { 001 / 003 / 021 / 201 }	(-B₃)	(-B₂)	(-B₁)	(-A)	(+B₃)	(+B₂)	(+B₁)	(+A)

FIG. 2. Representation of $T^{\phi\Gamma}$ tensors in the Cartesian basis $H_{\hat{n}}^{\Gamma}$ for $N_M = 3$.

zero. We will notice that each time there exists only a tensor which respects the selection rules.

Given these elements, it becomes elementary to decompose the Boltzmann equation according to the invariant subspaces of $(S_2 \times V)$. Especially, the Chapman—Enskog process[1] can be developed by using the reducibility of the collision operator and by expanding the first member of the equation in irreducible tensors for the transformations of $(S_2 \times V)$. This calculation is straightforward and can be formally followed in a similar way to that used by Kumar for the SO_3 group.

We will not undertake this study because it falls outside the limited framework of this paper.

Let us point out, however, that this calculation involves, in the reduction of the first member of the Boltzmann equation, the evaluation of matrix elements of the tensors $\mathbf{v}$ and

$$\nabla_v = \left\{ \frac{\partial}{\partial v_x}, \frac{\partial}{\partial v_y}, \frac{\partial}{\partial v_z} \right\}.$$

For instance, for v_x we easily find with $\nu^2 = \beta m / 2$

$$\int W H_{\mathbf{n}_1}^{S_1} v_x H_{\mathbf{n}_2}^{S_2} d\mathbf{v} = \frac{1}{2\nu} \left(\sqrt{2(n_{2x} + 1)} \delta_{n_{1x}, n_{2x}+1} \right.$$
$$\left. + \sqrt{2n_x} \delta_{n_{1x}, n_{2x}-1} \right) \delta_{n_{1y} n_{2y}} \delta_{n_{1z} n_{2z}} \qquad (3.36)$$

and for ∂/∂_x,

$$W H_{\mathbf{n}_1}^{S_1} \frac{\partial}{\partial v_x} H_{\mathbf{n}_2}^{S_2} d\mathbf{v}$$

$$= \sqrt{2n_x} \nu \delta_{n_{1x}, n_{2x}-1} \delta_{n_{1y}, n_{2y}} \delta_{n_{1z}, n_{2z}}. \qquad (3.37)$$

Therefore the v_x and $\partial/\partial v_x$ operators are irreducible tensorial operators $T^{(-B_1)}$ and very generally the velocity $\mathbf{v}$ and the operator ∇ belong to the same invariant subspaces $(-B_i)$.

Finally, we have to note that all the results which we have established are still valid for a gas mixture for the structure of the Boltzmann equation and therefore the collision integral possesses the same properties with regard to the Klein V group, whatever the number of gas constituents. On a practical basis we operate in the $\mathcal{E} = \mathcal{E}_1 \times \mathcal{E}_2 \times \cdots \times \mathcal{E}_G$ tensorial product space of the G vectorial spaces corresponding to each gas; each subspace $\mathcal{E}^{\phi\Gamma}$ of $\mathcal{E}$ remains invariant but its dimensions are multiplied by the number of gas constituents.

4. STUDY OF D IN THE CYLINDRICAL BASIS C_{kmn_z}

A. Reducibility of D in this basis

In this section we shall very rapidly show the reducibility of the integral collision operator in an L^2 orthogonal basis of the same weight W as the basis constructed on the Hermite polynomials but which is naturally expanded with regard to cylindrical coordinates.

Let C_{kmn_z} be the basis whose explicit definition is given in Appendix A. The parameters corresponding to the velocity $\mathbf{v}$ are denoted by $\{V, \varphi, v_z\}$. We can write

$$C_{kmn_z}(V, \varphi, v_z) = U_{km}(V)(\exp(im\varphi)/\sqrt{2\pi}) H_{n_z}(v_z), \qquad (4.1)$$

where k, n_z are integers ≥ 0 and m is an integer.

In this basis the matrix elements of D are given by

$$\int C_{k'm\,n'_z}^* [C_{kmn_z}] d\mathbf{v}_1 = \langle k'm'n'_z | D | kmn_z \rangle = D_{k'm', km} \qquad (4.2)$$

with $\mathbf{k} = \{k, n_z\}$.

Let $\langle plm' | kmn_z \rangle$ be the unitary transformation coefficients from the basis $Q_{plm'}$ to the basis $\{C_{kmn_z}\}$. We have

$$C_{kmn_z} = \sum_{plm'} \langle plm' | kmn_z \rangle Q_{plm'}. \qquad (4.3)$$

With the selection rules [Appendix B, Eq. (A4)] the sum over m' in (4.3) disappears. The operator D being diagonal in l, m in the basis $|plm\rangle$, we finally have

$$D_{k'm', km} = \sum_{pp'lm''} \langle k'm'n'_z | p'lm'' \rangle D_{p',p}^l \langle plm'' | kmn_z \rangle. \qquad (4.4)$$

Moreover the $D_{k'm', km}$ matrix elements are zero except when

$$m' = m'' = m, \quad (-)^{n'_z} = (-)^{n_z} = S_z. \qquad (4.5)$$

Consequently, the D operator is diagonal in m and such that the parity of n_z is conserved. Thus

$$D_{k'm', km} = D_{(k'), (k)}^m \delta_{mm'}, \qquad (4.6)$$

where (k) is such that $(-)^{n_z} = S_z$. The operator D is therefore reducible into two subspaces labelled by the parity $S_z = \pm 1$ and in each subspace into a set of invariant subspaces identified by m.

B. Dimension of the invariant subspaces for a finite dimension basis

Let N_M be the maximum degree of the polynomial C_{kmn_z}. The dimensions of the submatrices $D_{(k'), (k)}^m$ are obtained by evaluating all the possible combinations of the values of k and n_z such as
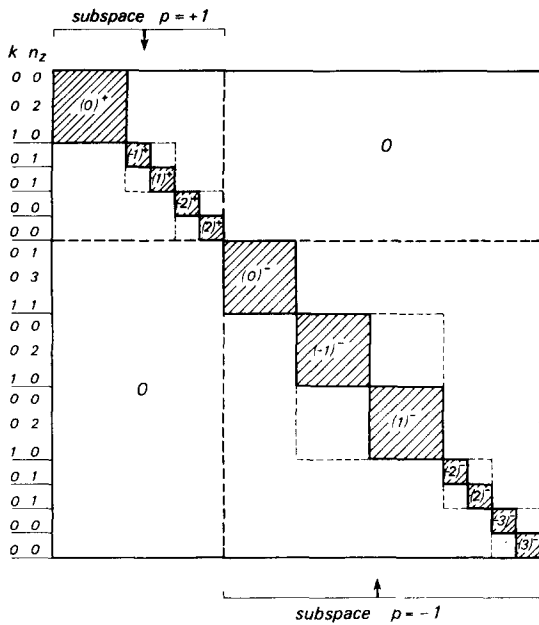
FIG. 3. Representation of the integral Boltzmann operator in the cylindrical basis $C_{(k)}^{pm}$ for $N_M = 3$. The value of the couple (pm) is noticed in the figure by $(m)^p$ and moreover $(\bar{k}) = (k\,n_z)$.

$$2k + n_z + |m| \leqslant N_M, \qquad (4.7)$$

with $(-)^{n_z} = S_z$.

We notice immediately that the matrices $D^m_{(\bar{k}')(\bar{k})}$ and $D^{-m}_{(\bar{k}')(\bar{k})}$ have the same dimensions, since the relation (4.7) only depends on $|m|$.

After some calculations, we find

$$\dim[D^{\pm m}] = \frac{(K+1)(K+2)}{2}, \qquad (4.8)$$

with

$$K = \left[\frac{N_M - |m|}{2}\right] \quad \text{if } (-)^{n_z} = +1, \qquad (4.9)$$

and

$$K = \left[\frac{N_M - |m| - 1}{2}\right] \quad \text{if } (-)^{n_z} = -1. \qquad (4.10)$$

As an example, we give in Fig. 3 the structure of D with $N_M = 3$ (basis of 20 polynomials). However the state classification can only be understood with the support of the following.

C. Groups leading to the reducibility of D

These groups are the S_2 group generated by the parity operator P and the SO_2 group of the $R_z(\alpha)$ rotations by any angle α about the axis z. Given that P commutes with rotation operator R and that D commutes with P and with R, then with R_z, it is therefore possible to form an eigenbasis common to P, R_z, and D. We shall see that the basis $C_{kmn_z}(V, \varphi, v_z)$ is such a basis.

In a space inversion φ is changed into $(\varphi + \pi)$ and the axis z into $-z$. Using (4.1) and (3.3), we have

$$PC_{kmn_z}(V, \varphi, v_z)$$
$$= C_{kmn_z}(V, \varphi + \pi, -v_z) = (-)^{m+n_z}C_{kmn_z}(V, \varphi, v_z). \qquad (4.11)$$

Thus

$$P|kmn_z\rangle = (-)^{m+n_z}|kmn_z\rangle. \qquad (4.12)$$

In the same way, the $R_z(\alpha)$ rotation changes φ into $(\varphi + \alpha)$. Then

$$R_z(\alpha)C_{kmn_z}(V, \varphi, v_z) = C_{kmn_z}(V, \varphi + \alpha, v_z)$$
$$= \exp(im\,\alpha)C_{kmn_z}(V, \varphi, v_z) \qquad (4.13)$$

whence

$$R_z(\alpha)|kmn_z\rangle = \exp(im\,\alpha)|kmn_z\rangle. \qquad (4.14)$$

This result is well known for $\exp(im\,\varphi)$ forms a basis for the irreducible representations $\exp(im\,\alpha)$ of the SO_2 Abelian group. In the present case, the $(S_2 \times SO_2)$ irreducible representations are labelled by $p = (-)^{m+n_z}$ and m, whence by noting

$$|kmn_z\rangle = |p, m, (\mathbf{k})\rangle \qquad (4.15)$$

with $\mathbf{k} = (k, n_z)$ we can write, by applying the Wigner—Eckart theorem to the D operator,

$$\langle p', m', (\mathbf{k}')|D|p, m, (\mathbf{k})\rangle = \delta_{pp'}\delta_{mm'}D^{pm}_{(\mathbf{k}')(\mathbf{k})}. \qquad (4.16)$$

The SO_2 group being a continuous one, the number of its irreducible representations is not finite. Then it is the same with the invariant subspaces in which D is developed. However, in any subspace of finite dimension N_M, we have $m \in [-N_M, +N_M]$ according to (4.7) which finally gives $2(2N_M + 1)$ invariant subspaces with the two possible values for p. These are the subspaces seen in Fig. 3, each of them being labelled by the notation $m^p = \pm |m|^{\pm}$.

D. Equality of the matrices D^{pm} and D^{p-m}

As D commutes with any rotation operator, we have

$$D = R_x^{-1}(\pi)DR_x(\pi) = R_x^+(\pi)DR_x(\pi), \qquad (4.17)$$

where $R_x(\pi)$ is a rotation by $+\pi$ about the axis x. In this rotation, the axis z is changed into $-z$ and the angle φ into $-\varphi$. So,

$$R_x(\pi)|kmn_z\rangle = (-)^{n_z}|k - mn_z\rangle. \qquad (4.18)$$

By using (4.16), (4.17), and (4.18), we have

$$D^{pm}_{(\mathbf{k}')(\mathbf{k})} = (-)^{n_z+n_z'}D^{p-m}_{(\mathbf{k}')(\mathbf{k})}, \qquad (4.19)$$

but the conservation of the parity $p = (-)^{m+n_z}$ implies that $(-)^{n_z'+n_z} = 1$; therefore, finally

$$D^{pm} = D^{p-m}. \qquad (4.20)$$

The matrices D^{pm} representing D in the basis $|kmn_z\rangle$ are therefore only dependent on $|m|$.

E. Case of the nonlinear collision operator

This problem is dealt with in the same way as in Sec. 3. The expression (3.23) is evaluated in the coupled basis by introducing the Clebsch—Gordan coefficients $\langle p_1m_1, p_2m_2|pm\rangle$ of $(S_2 \times SO_2)$ such as

$$C^{pm}((\mathbf{k}_1)p_1m_1, (\mathbf{k}_2)p_2m_2)$$
$$= \langle p_1m_1, p_2m_2|pm\rangle C^{p_1m_1}_{(\mathbf{k}_1)} C^{p_2m_2}_{(\mathbf{k}_2)}. \qquad (4.21)$$

Examination of the characters of SO_2 and $S_2^{12,13}$ shows that we necessarily have

$$\langle p_1 m_1, p_2 m_2 | pm \rangle = \delta_{p_1 p_2, p} \delta_{m_1 + m_2, m} \ . \qquad (4.22)$$

Taking (4.21) into account, we can evaluate (3.24) in the coupled basis which gives

$$\int C^{pm*}(\alpha) D C^{p'm'}(\alpha') \, dv_2 = \langle \alpha, pm \, | D | \, \alpha', p'm' \rangle$$

$$= \delta_{pp'} \delta_{mm'} D_{\alpha\alpha'}^{pm} \, , \qquad (4.23)$$

where we applied the Wigner—Eckart theorem and we denoted by α and α' the index sets $[(k)pm, (0)1,0]$ and $[(k_1)p_1 m_1, (k_2)p_2 m_2]$.

The invariant subspaces are therefore the same for the operator D in the coupled basis as when D is linearized and evaluated in the simple basis $C_{(k)}^{pm}$.

The observations, concerning the decomposition in irreducible parts of the distribution function $f(\mathbf{r}, \mathbf{v}, t)$ or the tensorial operators T, are still applicable here. We have only to take the relations (3.31)—(3.35), replacing the indices $(p\Gamma)$ by (pm) and the Clebsch—Gordan coefficients $\langle p_1 \Gamma_1, p_2 \Gamma_2 | p\Gamma \rangle$ by $\langle p_1 m_1, p_2 m_2 | pm \rangle$.

Finally, the invariance under rotation of D being independent on the number of constituents of the gas, it is clear that D will still decompose following the same invariant subspaces whatever the number G of these constituents, the dimension of each subspace being multiplied by G.

5. REDUCIBILITY OF THE BOLTZMANN EQUATION IN THE PHASE SPACE $\mathcal{E}_r \times \mathcal{E}_v$

A. Proof

The Boltzmann equation can be expressed in the general form

$$\hat{D}[f] = D[f], \qquad (5.1)$$

where the operator $\hat{D}$ is

$$\hat{D} = \frac{\partial}{\partial t} + \mathbf{v} \cdot \nabla_r + \frac{\mathbf{F}}{m} \cdot \nabla_v \qquad (5.2)$$

and where D is given by the relation (3.23). The distribution function $f(\mathbf{r}, \mathbf{v}, t)$, solution of (5.1), is therefore defined in the space $\mathcal{E}_{rv} = \mathcal{E}_r \times \mathcal{E}_v$ of the $\mathbf{r}$ and $\mathbf{v}$ variables. The traditional methods to determine $f(\mathbf{r}, \mathbf{v}, t)$ are essential related to a decomposition of the distribution function in the space $\mathcal{E}_v$, independently of the properties that $\hat{D}$ and D possess with regard to $\mathcal{E}_r$.

We shall demonstrate that the consideration of properties of $\hat{D}$ and D in relation to the phase space $\mathcal{E}_{rv}$ gives us important information for the determination of the solutions f.

Let us consider first of all the case where the external forces $\mathbf{F}$ are zero. The operator $\hat{D}$ is simplified in $\hat{d}$, such as

$$\hat{d} = \frac{\partial}{\partial t} + v \cdot \nabla_r. \qquad (5.3)$$

The Boltzmann equation is written

$$\hat{d}[f] = D[f]. \qquad (5.4)$$

The collision operator D is independent by construction of the variable $\mathbf{r}$. It commutes with all the transformation operators relating to $\mathbf{r}$ and, in particular, with rotation operators and space inversion operators. It therefore commutes with all the operators of the O_3 group relating to $\mathbf{r}$. On the other hand, D still commutes with the operators of this same group when the transformations are applied on the variable velocity $\mathbf{v}$. Only this last result has been exploited by Kumar in the spherical basis and by ourselves in the Cartesian and cylindrical bases. Finally, D commutes with all the operators of the $[(O_3)_r \times (O_3)_v]$ group, the direct product of the transformations on $\mathbf{r}$ and $\mathbf{v}$.

If we consider the operator $\hat{d}$, we detect that this operator commutes with all the operations of the O_3 group (rotation plus space inversion) performed in the phase space. This is clear for the differential operator $\partial/\partial t$ independent of $\mathbf{r}$ and $\mathbf{v}$. On the other hand, the operator $\mathbf{v} \cdot \nabla_r$ is invariant for all space rotation and inversion performed simultaneously in $\mathcal{E}_v$ and $\mathcal{E}_r$.

In these circumstances, $\hat{d}$ commutes with all the operations of the group $(O_3)_{rv}$. As D commutes with the operations of $[(O_3)_r \times (O_3)_v]$, it commutes a fortiori with those of $(O_3)_{rv}$, for we have $[(O_3)_r \times (O_3)_v] \supset (O_3)_{rv}$.

Consequently, if we expand f on the irreducible representations of $(O_3)_{rv}$, the Wigner—Eckart theorem will be applicable to the two members of the Boltzmann equation and this will be totally resolved in subsequent independent equations, each one relating to determined invariant subspace.

The above observations can be used when the external forces $\mathbf{F}(\mathbf{r}, t)$ are no longer zero but possess a determined symmetry. If $\mathbf{F}$ is invariant under rotation about a fixed point O, the invariance group is still $(O_3)_{rv}$, the origin of frames being in O, for $((\mathbf{F}/m) \cdot \nabla_v)$ is an invariant of this group.

If $\mathbf{F}$ has an axis of determined symmetry, the invariance group of $((\mathbf{F}/m) \cdot \nabla_v)$ is then $(O_2)_{rv}$. As we have $(O_3)_{rv} \supset (O_2)_{rv}$, the Boltzmann equation is completely resolvable into the irreducible representations of $(O_2)_{rv}$. Then an explicit calculation requires the use of the cylindrical basis C_{kmn_z} which we discussed in Sec. 4.

Finally in the case where $\mathbf{F}$ has Cartesian symmetries, the invariance group of $((\mathbf{F}/m) \cdot \nabla_v)$ is the $(V)_{rv}$ Klein group. There again, $(O_3)_{rv}$ contains $(V)_{rv}$. In these circumstances, $\hat{D}$ and D are resolvable into the irreducible representations of V discussed in Secs. 2 and 3. The basis to use is clearly the Hermite polynomial H_n basis.

The construction of the irreducible representations of $(O_3)_{rv}, (O_2)_{rv}$, or $(V)_{rv}$ when the distribution function $f(\mathbf{r}, \mathbf{v}, t)$ is expanded on a local basis, is a difficult problem, because the parameters β and mean velocity $\mathbf{u}$, which are included in the definition of the weight function W, depend themselves on $\mathbf{r}$ and t and this dependence complicates the expression of its expansion. On the other hand this study is simple when the adopted basis is such that β and $\mathbf{u}$ are independent of $\mathbf{r}$ and t. As an example, we shall look at this in the next section

and then see the problem concerning the local and absolute basis.

B. Study of a particular case

In order to illustrate the technique to be used we are going to consider the simple case where $f(\mathbf{r},\mathbf{v},t)$ is expanded on an absolute basis ($\beta = \beta_0$ and $\mathbf{u} = \mathbf{u}_0$) by taking $\mathbf{F} = 0$. The cases where $\mathbf{F}$ is different from zero or is of determined symmetry may be dealt with similarly.

In this case, the invariant group is $(O_3)_{\mathbf{r}\mathbf{v}}$. The distribution function $f(\mathbf{r},\mathbf{v},t)$ is thus expanded on the $Q_{nlm}(\mathbf{v})$ basis,

$$f(\mathbf{r},\mathbf{v},t) = W \sum_{lmn} F_{nlm}(\mathbf{r},t) Q_{nlm}(\mathbf{v}). \tag{5.5}$$

According to the Chapman—Enskog process, the Boltzmann equation for a simple gas is written (Kumar,[5] relation 111) to the $(r-1)$ order approximation,

$$\sum_{n'l'm'} \langle nlm|\hat{d}|n'l'm'\rangle \,{}^{(r-1)}F_{n'l'm'}(\mathbf{r},t) + \sum_{s=1}^{r-1}$$

$$\times \sum_{\substack{n_1l_1m_1 \\ n_2l_2m_2}} \langle nlm|J|n_2l_2m_2, n_1l_1m_1\rangle$$

$$\times {}^{(r-s)}F_{n_2l_2m_2}(\mathbf{r},t)\,{}^{(s)}F_{n_1l_1m_1}(\mathbf{r},t)$$

$$= -2\sum_{\substack{n_1l_1m_1 \\ n_2l_2m_2}} \langle nlm|J|n_2l_2m_2, n_1l_1m_1\rangle$$

$$\times {}^{(0)}F_{n_2l_2m_2}(\mathbf{r},t)\,{}^{(r)}F_{n_1l_1m_1}(\mathbf{r},t), \tag{5.6}$$

where the matrix element $\langle nlm|J|n_1l_1m_1, n_2l_2m_2\rangle$ is given by Kumar[5] (relation 91).

Then, if we introduce the standard basis $T_M^L(nl\,l_r;\hat{r}\mathbf{v})$ of $(O_3)_{\mathbf{r}\mathbf{v}}$ such as

$$T_M^L(nl\,l_r;\hat{r}\mathbf{v}) = \sum_{m_r m} \langle lm\,l_r m_r|LM\rangle Y_{l_r m_r}(\hat{r}) Q_{nlm}(\mathbf{v}),$$

$$\hat{r} = \mathbf{r}/r, \tag{5.7}$$

denoting $\alpha = (nl\,l_r)$, we obtain

$$f(\mathbf{r},\mathbf{v},t) = W \sum_{LM\alpha} F_M^L(\alpha;rt) T_M^L(\alpha;\hat{r}\mathbf{v}). \tag{5.8}$$

Then with

$${}^{(s_1 s_2)}G_{M'}^{L'}(\alpha_1\alpha_2 L_1 L_2, rt)$$

$$= \sum_{M_1 M_2} \langle L'M'|L_1 M_1 L_2 M_2\rangle\,{}^{(s_1)}F_{M_1}^{L_1}(\alpha_1 rt)\,{}^{(s_2)}F_{M_2}^{L_2}(\alpha_2 rt) \tag{5.9}$$

the Boltzmann equation is written

$$\sum_{\alpha' L'M'} \langle \alpha LM|\hat{d}|\alpha'L'M'\rangle\,{}^{(r-1)}F_{M'}^{L'}(\alpha',rt) + \sum_{s=1}^{r-1} \sum_{\substack{\alpha_1\alpha_2 L_1 L_2 \\ L'M'}}$$

$$\langle \alpha LM|J|\alpha_1\alpha_2 L_1 L_2 L'M'\rangle\,{}^{(s,r-s)}G_{M'}^{L'}(\alpha_1\alpha_2 L_1 L_2;rt)$$

$$= -2 \sum_{\substack{\alpha_1\alpha_2 L_1 L_2 \\ L'M'}} \langle \alpha LM|J|\alpha_1\alpha_2 L_1 L_2 L'M'\rangle$$

$$\times {}^{(0,r)}G_{M'}^{L'}(\alpha_1\alpha_2 L_1 L_2;rt), \tag{5.10}$$

where

$$\langle \alpha LM|J|\alpha_1\alpha_2 L_1 L_2 L'M'\rangle = \sum_{M_1 M_2} \langle L_1 M_1 L_2 M_2|L'M'\rangle$$

$$\times \langle \alpha LM|J|\alpha_1 L_1 M_1, \alpha_2 L_2 M_2\rangle. \tag{5.11}$$

Under these circumstances, the Wigner—Eckart theorem can be applied and we obtain

$$\langle \alpha LM|\hat{d}|\alpha'L'M'\rangle = \delta_{LL'}\delta_{MM'}\hat{d}_{\alpha\alpha'}^L. \tag{5.12}$$

$$\langle \alpha LM|J|\alpha_1\alpha_2 L_1 L_2 L'M'\rangle = \delta_{LL'}\delta_{MM'}J_{\alpha,(\alpha_1\alpha_2 L_1 L_2)}^L. \tag{5.13}$$

Then, we have the system of equations

$$\sum_{\alpha'} \hat{d}_{\alpha\alpha'}^L\,{}^{(r-1)}F_M^L(\alpha',rt) + \sum_{s=1}^{r-1} \sum_{\alpha_1\alpha_2 L_1 L_2} J_{\alpha,(\alpha_1\alpha_2 L_1 L_2)}^L$$

$$\times {}^{(s,r-s)}G_M^L(\alpha_1\alpha_2 L_1 L_2, rt)$$

$$= -2 \sum_{\alpha\,\alpha_2 L_1 L_2} J_{\alpha,(\alpha_1\alpha_2 L_1 L_2)}^L\,{}^{(0,r)}G_M^L(\alpha_1\alpha_2 L_1 L_2;rt), \quad L \in \mathbb{N}. \tag{5.14}$$

Each relation is a differential equation which, in opposition to (5.6), is only relative to the variables r and t. On the other hand, the solution ${}^{(0,r)}G_M^L$ does not depend on M. It suffices therefore to calculate ${}^{(0,r)}G_M^L$ ($M \in [-L, +L]$) whence we deduce ${}^{(r)}F_M^L(\alpha,rt)$ by

$${}^{(0)}F_{M_1}^{L_1}(\alpha_1;rt)\,{}^{(r)}F_{M_2}^{L_2}(\alpha_2;rt)$$

$$= \sum_{LM} \langle LM|L_1 M_1 L_2 M_2\rangle\,{}^{(0,r)}G_M^L(\alpha_1\alpha_2 L_1 L_2;rt). \tag{5.15}$$

At last the functions ${}^{(0)}F_M^L(\alpha;rt)$ are given by the zero-order approximation of the distribution function, i.e.,

$$f^{(0)}(\mathbf{r},\mathbf{v},t) = n(\mathbf{r},t)W = W \sum_{LM\alpha} {}^{(0)}F_M^L(\alpha;rt) T_M^L(\alpha;\hat{r}\mathbf{v}),$$

$$\tag{5.16}$$

whence

$${}^{(0)}F_M^L(\alpha;rt) = \int Wn(\mathbf{r},t) T_M^{L*}(\alpha;\hat{r}\mathbf{v})\,d\hat{r}\,d\mathbf{v}$$

$$= \delta_{n0}\delta_{l0}\delta_{l_r L} \int n(\mathbf{r},t) Y_{LM}^*(\hat{r})\,d\hat{r}. \tag{5.17}$$

Finally, we establish that there exists a Chapman—Enskog hierarchy relative to the value of L and these hierarchies are independent.

The Eq. (5.14) therefore represents the most complete expansion of the Boltzmann equation in the absence of external forces ($\mathbf{F} = 0$).

If the external forces are no longer zero but have a symmetry center O, we situate the origin of the frame of reference in this point, which gives the same result (5.14), with only the matrix element $\hat{d}_{\alpha\alpha'}^L$ being modified. If $\mathbf{F}$ has a cylindrical or Cartesian symmetry, we use the same method in order to expand $f(\mathbf{r},\mathbf{v},t)$ on the basis introduced in Secs. 4 and 3.

Let us notice here that the expansion (5.14) was obtained with the hypothesis β and $\mathbf{u}$ being independent of $\mathbf{r}$ and t. If this was not so, the expansion would have no

incidence on a theoretical point of view, because the result (5.14) is obtained from the invariance of the operators $\hat{D}$ and D under the transformations of various groups. This invariance is independent of whether the basis is absolute or local.

However, the use of a local basis presents a certain technical complexity because β and $\mathbf{u}$ must be expanded on the spherical harmonics $Y_{lm}(\hat{r})$. This complexity therefore decreases the interest of the result (5.14) when the basis is local. However, it would seem clear that a local basis is generally better adapted to the description of the physical phenomena than an absolute basis. This is because the variations of temperature β and of mean velocity $\mathbf{u}$ are also included in the weight function W and not only in the $F_{nlm}(\mathbf{r}, t)$ coefficients, as in an absolute basis. We shall see, however in the following paragraph that each moment of the distribution function, expanded on a local basis, is a finite linear combination of the moments of this same function, expanded on an absolute basis. The two bases are therefore equivalent in any L^2 subspace of finite dimension for the determination of the distribution function moments.

C. Equivalence of "absolute" and "local" moments in any subspace of finite dimension

This equivalence exists independently of the spherical, Cartesian, or cylindrical choice of the basis. We make calculations in the usual spherical basis $\{Q_{nlm}\}$ and the parameters of the absolute basis will be noted ν_0 $= \sqrt{\beta_0 m/2}$ and $\mathbf{u}_0$, ν and $\mathbf{u}$ being those of the local basis. In order to simplify the calculations, we shall first consider the case where $\mathbf{u} = \mathbf{u}_0$.

So, we have

$$f(\mathbf{r}, \mathbf{v}, t) = \sum_{nlm} W^0 F^0_{nlm}(\mathbf{r}, t) Q_{nlm}(\nu_0 \mathbf{v}) = \sum_{nlm} W F_{nlm}(\mathbf{r}, t) Q_{nlm}(\nu \mathbf{v}).$$

$$(5.18)$$

But with the definition of Q_{nlm} (Appendix A) we can see that $Q_{nlm}(\nu \mathbf{v})$ and $Q_{nlm}(\nu_0 \mathbf{v})$ are orthogonal polynomials with respect to the $\hat{v}$ variable. Then we find

$$F_{nlm}(\mathbf{r}, t) = \int Q^*_{nlm} f \, d\mathbf{v}$$

$$= \sum_{n'l'm'} F^0_{n'l'm'} \int W^0 Q^*_{nlm}(\nu \mathbf{v}) Q_{n'l'm'}(\nu_0 \mathbf{v}) \, d\mathbf{v}$$

$$= \sum_{n'} F^0_{n'lm} \int W^0 R_{nl}(\nu v) R_{n'l}(\nu_0 v) v^2 \, dv, \qquad (5.19)$$

but $R_{nl}(\nu, v)$ is a polynomial in v of degree $(2n + l)$. We can therefore expand it in a finite linear combination of the polynomials $R_{n'l}(\nu_0 v)$ and we obtain

$$R_{nl}(\nu v) = \sum_{n'0}^{n} C^n_{n'}(l, \nu/\nu_0) R_{n'l}(\nu_0 v), \qquad (5.20)$$

where $C^n_{n'}$ is given in Appendix C.

From this, and by using the orthogonality of the polynomials $R_{nl}(\nu_0 v)$ we deduce

$$F_{nlm}(\mathbf{r}, t) = \sum_{n'=0}^{n} C^n_{n'}(l, \nu/\nu_0) F^0_{n'lm}(\mathbf{r}, t). \qquad (5.21)$$

Thus the moments $F_{nlm}(\mathbf{r}, t)$ of the local basis are

finite linear combinations of the moments $F^0_{nlm}(\mathbf{r}, t)$ of the absolute basis. The converse is evident. Finally, we must observe that the n first moments F^0_{nlm} enable us to calculate the n corresponding F_{nlm} moments.

The interest of the relation (5.21) is to allow the utilization of an absolute basis for the solution of the Boltzmann equation and therefore finally the exploitation in particular of the relation (5.14).

This relation was established by making the hypothesis $\mathbf{u} = \mathbf{u}_0$. The result, however, was still acquired in the general case because it is a consequence of the fact that the set Q_{nlm} is a polynomial set and that it is always possible to expand $Q_{nlm}(\nu(\mathbf{v} - \mathbf{u}))$ on the set $\{Q_{n'l'm'}(\nu_0(\mathbf{v} - \mathbf{u}_0))\}$.

The demonstration can also be made simply by using the Cartesian basis $\{H^{\phi\Gamma}_{\mathbf{a}}\}$. So we have

$$f(\mathbf{r}, \mathbf{v}, t) = \sum_{\mathbf{a}} W^0 F^{0(\phi\Gamma)}_{\mathbf{a}} H^{\phi\Gamma}_{\mathbf{a}}(\nu_0(\mathbf{v} - \mathbf{u}_0)) \qquad (5.22)$$

$$= \sum_{\mathbf{a}} W F^{(\phi\Gamma)}_{\mathbf{a}} H^{\phi\Gamma}_{\mathbf{a}}(\nu(\mathbf{v} - \mathbf{u})) \qquad (5.23)$$

and by following the same process as before, we finally obtain the result

$$F^{(\phi\Gamma)}_{\mathbf{a}}(\mathbf{r}, t) = \sum_{\mathbf{a}'=0}^{\mathbf{a}} K^{\mathbf{a}}_{\mathbf{a}'}(\nu/\nu_0, \mathbf{u}_0 - \mathbf{u}) F^{0(\phi\Gamma)}_{\mathbf{a}'}(\mathbf{r}, t), \qquad (5.24)$$

where $K^{\mathbf{a}}_{\mathbf{a}'}$ is given in Appendix C.

NOTE: The results of Sec. 5, Parts A and B were established in group terms on $(\mathcal{E}_r \times \mathcal{E}_v)$. We also can express them in an equivalent way in algebraic terms: The reducibility of the Boltzmann equation in the subspace $\mathcal{E}_{\hat{r}} \times \mathcal{E}_{\hat{v}}$ of $\mathcal{E}_r \times \mathcal{E}_v$ shows that the basis used for this reduction is an eigenbasis of the Boltzmann equation in this subspace. Explicitly, when for example the external forces are zero or invariant under rotation about a fixed point, this basis is given by the examination of the relations (5.7) and (5.9), i.e., finally for the subspace $(\mathcal{E}_{\hat{r}} \times \mathcal{E}_{\hat{v}})$ by Ψ^L_M such as

$$\Psi^L_M(l_1 l'_1 L_1, l_2 l'_2 L_2; \hat{r}, \hat{v}_1, \hat{v}_2)$$

$$= \sum_{\substack{M_1 M_2 m_1 m'_1 \\ m'_2 m'_2}} \langle L_1 M_1 L_2 M_2 | LM \rangle$$

$$\times \langle l_1 m_1 l'_1 m'_1 | L_1 M_1 \rangle \langle l_2 m_2 l'_2 m'_2 | L_2 M_2 \rangle$$

$$\times Y_{l_1 m_1}(\hat{v}_1) Y_{l'_1 m'_1}(\hat{r}) Y_{l_2 m_2}(\hat{v}_2) Y_{l'_2 m'_2}(\hat{r}). \qquad (5.25)$$

In the cylindrical and Cartesian basis, expressions of the same type are established by considering the functions $H^{\phi\Gamma}_{(\mathbf{a})}$ and $C^{\phi m}_{(\mathbf{k})}$ as well as the corresponding Clebsch–Gordan coefficients.

6. CONCLUSION

The results obtained in the above sections show the interest presented by the theorems of the group theory to solve practical problems concerning the structure of the Boltzmann equation, and this without passing by the intermediary of complicated algebraic equations. We have thus been able to show that the Boltzmann collision operator was reduced in the determined basis such as

$\{H_{\mathbf{n}}\}$ or $\{C_{kmn_z}\}$ in a direct sum of matrices, each of them belonging to invariant subspaces. The groups involved in this reduction have been identified (V) and SO_2 groups) and the invariance of the collision operator under the rotations was fully exploited to prove the equality of different matrices of the expansion following specific rules.

The above results have been enlarged to the full Boltzmann equation. This has enabled us to define totally the structure of this equation in the space ($\mathcal{E}_{\mathbf{r}} \times \mathcal{E}_{\mathbf{v}}$). Full application of the resulting properties has finally been made in the case where the external force **F** is zero or invariant under rotation.

This application enabled us to find the eigenfunctions of the Boltzmann equation in the subspace ($\mathcal{E}_{\hat{r}} \times \mathcal{E}_{\hat{v}}$) and to show how the Chapman—Enskog process could be used. The method followed in this particular case can be utilized point by point when the force **F** is of cylindrical or Cartesian symmetry. It gives very similar results to the previous methods if the Clebsch—Gordan coefficients for the corresponding groups are used together with the indices (pm) or $(p\Gamma)$ of their irreducible representations.

The previous considerations lead to a great simplification in the determination of the distribution function $f(\mathbf{r}, \mathbf{v}, t)$, for instance in the cases of the Chapman—Enskog procedure or the direct calculation process for the linearized equation.

It must also be noticed that the research and utilization of the Boltzmann equation symmetries allow us to derive the selection rules, which in an essential manner depend on the boundary conditions of the problem, to which $f(\mathbf{r}, \mathbf{v}, t)$ is subjected through its expansion on different basis. The knowledge of these selection rules is a fundamental matter since they directly lead to the reduction of theoretical as well as numerical calculations.

ACKNOWLEDGMENT

APPENDIX A

We give without commentary the definitions of the various functions used.

(a) *Spherical basis* $\{Q_{nlm}\}$:

$$Q_{nlm}(\nu\mathbf{v}) = R_{nl}(\nu v) Y_{lm}(\hat{v}) \quad (\hat{v} = \mathbf{v}/v),$$

where $Y_{lm}(\hat{v})$ is the usual spherical harmonic (see Messiah[8] for example) and

$$R_{nl}(\nu v) = \frac{1}{\sqrt{N_{nl}}} (\nu v)^l L_n^{l+1/2}(\nu^2 v^2),$$

$$L_n^{l+1/2}(x) = \sum_{k=0}^n (-)^k 2^k \binom{n}{k} \frac{(2l+1)!!}{(2l+2k+1)!!} x^k,$$

$$N_{nl} = \frac{n! [(2l+1)!!]^2}{2^{l-n+2}(2n+2l+1)!!\pi},$$

$$\int W Q_{nlm}^* Q_{n'l'm} \, d\mathbf{v} = \delta_{nn'} \delta_{ll'} \delta_{mm'},$$

with

$$W = \frac{\nu^3}{\pi^{3/2}} \exp(-\nu^2 v^2).$$

(b) *Cartesian basis* $H_{\mathbf{n}}$:

$$H_{\mathbf{n}}(\nu\mathbf{v}) = H_{n_x}(\nu v_x) H_{n_y}(\nu v_y) H_{n_z}(\nu v_z), \quad \mathbf{n} = (n_x, n_y, n_z),$$

$$H_{\mathbf{n}}(\nu\mathbf{v}) = \frac{1}{(2^n n!)^{1/2}} \sum_{p=0}^{[n/2]} (-)^p 2^{n-p} \binom{n}{2p} (2p-1)!!(\nu v)^{n-2p},$$

$$\int W H_{\mathbf{n}}(\nu\mathbf{v}) H_{\mathbf{n}'}(\nu\mathbf{v}) \, d\mathbf{v} = \delta_{\mathbf{nn}'}.$$

(c) *Cylindrical basis* $\{C_{kmn_z}\}$:

$$C_{kmn_z}(\nu\mathbf{v}) = U_{km}(\nu v)(\exp(im\varphi)/\sqrt{2\pi}) H_{n_z}(\nu v_z)$$

with

$$U_{km}(\nu V) = \left(\frac{2\pi k!}{[(k+m)!]^3}\right)^{1/2} (\nu V)^m L_k^m(\nu^2 v^2)$$

and

$$\int W C_{kmn_z}(\nu\mathbf{v}) C_{k'm'n_z'}^*(\nu\mathbf{v}) \, d\mathbf{v} = \delta_{kk'} \delta_{mm'} \delta_{n_z n_z'}.$$

APPENDIX B

A. Unitary transformation coefficients $\langle nlm \mid n_x n_y n_z \rangle$ of $(Q_{nlm} \leftrightarrow H_{\mathbf{n}})$

These coefficients have been examined by several authors. Their explicit value is given by Domergue[11] to which we refer for the calculations. We have for $m \geq 0$,

$$\langle nlm \mid n_x n_y n_z \rangle = \left(\frac{N_{nL}}{N_{\mathbf{n}}}\right)^{1/2} i^{n_y} \frac{(-)^{m+m+n_y}}{(2l+1)!!}$$

$$\times \left(4\pi(2l+1)\frac{(l-m)!}{(l+m)!}\right)^{1/2} \frac{n_x! n_y! n_z!}{n!} \left(\frac{n_x + n_y + m}{2}\right)!$$

$$\times \sum_s (-)^s \frac{(2l-2S)!(n+S)!}{S!(l-S)!(l-m-2S)![S-(l-m-n_z)/2]}$$

$$\times \sum_t (-)^t \left\{t!(n_y-t)!\left(t - \frac{n_y - n_x - m}{2}\right)!\right.$$

$$\left. \times [(n_x + n_y - m)/2 - t]!\right\}^{-1}$$

with $N_n = 2^{n_x + n_y + n_z} n_x! n_y! n_z!$, and the following rules:

$$\langle n_x n_y n_z \mid nlm \rangle = \langle nlm \mid n_x n_y n_z \rangle^* = (-)^{n_y} \langle nlm \mid n_x n_y n_z \rangle,$$

$$\langle nl - m \mid n_x n_y n_z \rangle = (-)^{m+n_y} \langle nlm \mid n_x n_y n_z \rangle, \tag{B1}$$

$$2n + l = n_x + n_y + n_z,$$

$$(-)^{n_x + n_y + m} = +1, \tag{B2}$$

$$(-)^{n_z + l + m} = +1.$$

B. Unitary transformation coefficients $\langle nlm' \mid kmn_z \rangle$ of $(Q_{nlm} \leftrightarrow C_{kmn_z})$

The calculation of these coefficients was carried out by Talman.[14] These are given by

$$\langle nlm \mid kmn_z \rangle = (-)^{(m+|m|)/2} \left(\frac{(l-|m|)!}{(l+|m|)!}\right)^{1/2} (-)^{k+n}$$

$$\times\left(\frac{n_z!\,(k+|m|)!\,(2l+1)}{k!\,n!\,(2l+2n+1)!\,!\,2^{n_z+l-n}}\right)^{1/2}$$

$$\times\sum_p\frac{(-)^p(2l-2p)!\,(p+n)!}{p!\,(l-p)!\,(2k-2n+n_z-2p)!\,(p+n-k)!}\ ,$$

and also with

$$\langle nlm\,|\,kmn_z\rangle=\langle kmn_z\,|\,nlm\rangle,$$

$$\langle nlm\,|\,km'n_z\rangle=\delta_{mm'}\langle nlm\,|\,kmn_z\rangle,\tag{B3}$$

$$(-)^{m+n_z+l}=+1,$$

$$2n+l=2k+n_z+|m|.\tag{B4}$$

APPENDIX C

Transformation coefficients of $R_{nl}(\nu v)$ and $H_{\mathbf{n}}^{p\,\Gamma}(\nu(\mathbf{v}-\mathbf{u}))$ polynomials follow:

$$C_{n'}^{n}\left(l,\frac{\nu}{\nu_0}\right)=\left(\frac{(2n+2l+1)!\,!\,(2n'+2l+1)!\,!}{2^{n+n'}\,n!\,n'!}\right)^{1/2}$$

$$\times\left(\frac{\nu}{\nu_0}\right)^l\sum_{k\,0}^{n}\sum_{k'\,0}^{n'}(-)^{k+k'}\binom{n}{k}\binom{n'}{k'}$$

$$\times\frac{(2k+2k'+2l+1)!\,!}{(2l+2k+1)!\,!\,(2l+2k'+1)!\,!}\left(\frac{\nu}{\nu_0}\right)^{2k},$$

$$K_{\mathbf{n}'}^{\mathbf{n}}=K_{n_x'}^{n_x}K_{n_y'}^{n_y}K_{n_z'}^{n_z}\ ,$$

$$K_{n'_x}^{n_x}=\sum_{x\,\,\,\,\,\,\,\,\,I\,0}^{[n_x/2]}\sum_{k\,n'_x}^{n_x-2p}(-)^I\,2^{n_x-2p-k}$$

$$\times\frac{n_x!}{l!\,n'_x!\,[(k-n'_x)/2]!\,(n_x-2l-k)!}$$

$$\times\left(\frac{\nu}{\nu_0}\right)^{n_x-2l}[\nu_0(u_0-u)]^{n_x-2l-k}.$$

*Also Institut de Physique Nucléaire, Université Claude Bernard (Lyon 1) 43, Boulevard du 11 Novembre 1918, 69621, Villeurbanne (France).

†Also Laboratoire de Spectrométrie Ionique et Moléculaire, Université Claude Bernard (Lyon 1) 43, Boulevard du 11 Novembre 1918, 69621, Villeurbanne (France).

[1]S. Chapman and T.G. Cowling, *The mathematical theory of non-uniform gases* (Cambridge U.P., Cambridge, 1952).

[2]J.O. Hirschfelder, C.F. Curtis, and R.B. Bird, *The molecular theory of gases and liquids* (Wiley, New York, 1954).

[3]J.H. Ferziger and H.G. Kaper, *Mathematical theory of transport processes in gases* (North-Holland, Amsterdam, 1972).

[4]K. Kumar, Ann. Phys. 37, 113 (1966).

[5]K. Kumar, Austr. J. Phys. 20, 205 (1967).

[6]K. Kumar, Austr. J. Phys. 23, 505 (1970).

[7]F. Reif, *Fundamentals of statistical and thermal physics* (McGraw-Hill, New York, 1965).

[8]A. Messiah, *Mécanique quantique* (Dunod, Paris, 1964).

[9]E.P. Wigner, *Group Theory* (Academic, New York, 1959).

[10]H. Grad, *Principles of the kinetic theory of gases, Handbuch der Physik* 12 (Springer, Berlin, 1960); Phys. Fluids 6, 147 (1963).

[11]L. Domergue, C.R. Acad. Sci. B 264, 403 (1967).

[12]H. Weyl, *The classical groups* (Princeton U.P., Princeton, N.J., 1946).

[13]M. Hamermesh, *Group theory and its application to physical problems* (Addison-Wesley, Reading, Mass., 1964).

[14]J.D. Talman, Nucl. Phys. A 141, 273 (1970).

1403   J. Math. Phys., Vol. 17, No. 8, August 1976

J.N. Massot and R. Bacis   1403

One-component plasma in $2+\epsilon$ dimensions

C. Deutsch

Laboratoire de Physique des Plasmas, Université Paris XI, 91405-Orsay, France
and Research Laboratory of Electronics, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139
(Received 9 February 1976)

The one-component plasma (ocp) model with neutralizing background is extended to real dimensionality $\nu = 2 + \epsilon$ with $-2 \le \epsilon \le 2$. The equilibrium properties (pair correlation and thermodynamic functions) investigated within the Debye approximation, up to the second-order in the plasma parameter $e^2/k_B T \lambda_D^\epsilon$, with the aid of the Wilson quadratures, interpolate between two- and three-dimensional results for $0 < \epsilon < 1$, and extend the $\nu = 3$ behavior to all $\nu \le 2$. The dimensionality $\nu = 2$ is shown to play a special role. Quantum diffraction corrections are included in the high temperature limit through a temperature-dependent effective Coulomb interaction. As a by-product, the particle diffusion coefficient (Bohm) of the strongly magnetized two-component plasma taken in the fluid limit may be given a finite volume-independent expression in the thermodynamic limit when $\nu = 2$, provided due attention is paid to the Tauberian properties of the Coulomb potential for $-2 \le \epsilon \le 0$.

I. INTRODUCTION

It is a common experience that the well-known one-component classical plasma model (ocp) with a neutralizing background exhibits significant modifications of its properties when the dimensionality parameter ν takes various integer values. This model has received a considerable attention, in the context of plasma physics as well as in the framework of basic equilibrium statistical mechanics. It is now a well-known fact[1] that, for $\nu \le 2$, the total Coulomb interaction energy per particle does not remain finite in the $V \to \infty$ limit, a phenomenon immediately explained by the analytical form of the ν-dimensional Coulomb interaction,

$$\phi^{(\nu)}(r) = \begin{cases} \text{sgn}(\nu - 2)|r|^{2-\nu}, & \nu \neq 2, \\ \ln|r|^{-1}, & \nu = 2. \end{cases} \quad (I.1)$$

Solution of the Poisson equation

$$\Delta\phi^{(\nu)}(r) = -|\nu - 2|S_\nu \delta_\nu(r), \quad S_\nu = \frac{2\pi^{\nu/2}}{\Gamma(\nu/2)}. \quad (I.2)$$

Then the total Coulomb energy of N particles with unit charge e, in the presence of an inert and homogeneous neutralizing background reads[1] $(r_{ij} = |r_i - r_j|)$

$$W^{(\nu)}(N) = \frac{e^2}{2} \sum_{i \neq j} \phi^{(\nu)}(r_{ij}) - \rho e^2 \sum_{j=1}^{N} \int d^\nu r$$

$$\times \phi^{(\nu)}(|r - r_j|) + \frac{\rho^2 e^2}{2} \iint d^\nu r \, d^\nu r'$$

$$\times \phi^{(\nu)}(|r - r'|) = \frac{e^2}{2} \sum_{i \neq j}^{N} \phi^{(\nu)}(r_{ij})$$

$$+ \frac{Ne^2}{2R^{\nu-2}} \sum_{i=1}^{N} \left(\frac{|r_i|}{R}\right)^2 - \tilde{B}_\nu(\nu, R, N), \quad \rho = \frac{N}{V}. \quad (I.3)$$

With the background self-energy $[V = (S_\nu/\nu)R^\nu]$

$$B_\nu(\nu, R, N) = \left(\frac{\nu+1}{\nu+2} + \phi^{(\nu)}(R)R^{\nu-2}\right)\frac{e^2 N^2}{2R^{\nu-2}}$$

$$= \left[|\nu - 2|\left(\frac{\nu+1}{\nu+2}\right) + \text{sgn}(\nu - 2)\right]\frac{e^2}{2}\left(\frac{\rho S_\nu}{\nu}\right)^{1-2/\nu}$$

$$\times N^{1+2/\nu}, \quad \nu \neq 2,$$

$$\left[\frac{\nu+1}{\nu+2} + \frac{1}{2}\ln\left(\frac{\rho S_\nu}{\nu N}\right)\right]\frac{e^2 N^\nu}{2}, \quad \nu = 2. \quad (I.4)$$

Equation (i.3) embodies all the previously considered ocp with integer ν. The case $\nu = 1$ with

$$W^{(1)}(N) = \frac{-e^2}{2}\sum_{i=1}^{N}(2i - 1 - N)r_i + e^2\rho\sum_{i=1}^{N}r_i^2 - e^2\rho R\sum_{i=1}^{N}r_i$$

$$+ \frac{e^2\rho NR^2}{2} - \frac{e^2\rho^2 R^3}{3} \quad (I.5)$$

has been extensively considered by Baxter and Kunz,[2] while the $\nu = 2$ ocp with

$$W^{(\nu)}(N) = \frac{e^2}{2}\sum_{i \neq j}^{N}\ln\left|\frac{r_{ij}}{R}\right| + \frac{Ne^2}{R}\sum_{i=1}^{N}\left(\frac{|X_i|^2}{R}\right)^2$$

$$- \left[\frac{3N^2}{4} + \frac{N}{4}\ln\left(\frac{\rho S_2}{2N}\right)\right]e^2 \quad (I.6)$$

has recently[1] received a lot of attention. As well as any other ocp, it provides a simple tool for testing the coherence of the basic principles of the Gibbsian ensemble theory for particles interacting through a long-range potential. Also, it allows for a very useful modelling of the strongly magnetized $\nu = 3$ plasma. Equation (I.3) shows that $\nu = 2$ is a landmark for the $\phi^{(\nu)}(\nu)$ long-range behavior. For $\nu \ge 2$, the absolute value of the latter is a decreasing function of r, while it increases with it for $\nu < 2$. Therefore, in order to render more systematic and more transparent the ν-dependence of the equilibrium properties (pair correlation and thermodynamic functions), and also to allow for a local study around $\nu = 2$, badly needed for the determination of the particle diffusion coefficient in a real two-component magnetized plasma,[3] we are led to extend (I.3) to any real ν. This process is an interpolation of the above integer-dimensionalized ocp models with an infinite number of real ones. Indeed, it is a very common technique both in atomic physics[4] and in phase transition theory,[5] as well as in high-energy physics[6] to realize (and even complexity[6]) the relevant parameter: space dimensionality or angular momentum to get additional information for the physical models with discrete values of their quantities. Generally speaking, it allows for the introduction of a much wider perspective through the unification and extension of apparently uncorrelated techniques and results. We expect the same kind of proce-

dure to be fruitful in the theory of the one-component plasma, where so many results have already been gathered in $\nu = 1, 2$, and 3 dimensions. Our feeling is supported by the inverse quadratic form of the Coulomb interaction in momentum space [Eq. (I. 1)]

$$\phi^{(\nu)}(k) = - S_\nu/k^2, \quad \text{all } \nu, \tag{I. 7}$$

allowing the extension of the Poisson equation to any ν values. A first by-product of the extension (I. 7), already noticed previously,[1] is the ν-independence of many features of the high-temperature (Debye) treatment in k-space of the equilibrium properties. As an example, the limits $g_2(0)$ and $g_2(\infty)$ of the pair correlation function exhibit analogous trends for all ν. However, the range of validity of the high-temperature approximation based upon a perturbative expansion with respect to the plasma parameter Λ remains strongly ν-dependent. The introduction of an infinite number of ocp models with a $r^{-(\nu-2)}$ interaction between $\nu = 2$ and 3 should give access to a selective extrapolation of the suitable $\nu = 2$ behaviors needed in the modelization of the three-dimensional magnetized plasma, while keeping away the pathological features arising from the nonvanishing harmonic term in the $V \to \infty$ limit for $\nu \leqslant 2$. This latter is responsible for the breakdown of the translational invariance used in the standard Debye analysis of the Mayer—Salpeter diagrams. Recent calculations[2] have revealed that the $\nu = 1$ ocp model exhibits a periodic (rather than fluidlike) equilibrium structure persisting to all temperatures. An important question is the determination of the smallest ν retaining the translational invariance, thus allowing for a continuous extrapolation of the $\nu = 3$ equilibrium properties. The Debye expansion performed below makes clear that $\nu_{\min} = 2 + \epsilon$ with $\epsilon > 0$. In this perspective, the interest of the real ν extension of the Coulomb interaction seems to lie in a more flexible modelization of the three-dimensional plasma, as explained below in Sec. IV where a well-defined Bohm-like particle diffusion coefficient is detained in the $\epsilon = 0$ limit.

On the other hand, the projection of the $\nu = 3$ ocp onto smaller ν ocp's could be of value for the determination of high order corrections to the usual model companion, free from the well-known short-range divergence of the r^{-1} interaction. Whenever possible, this procedure combined with a $\Delta\nu$ expansion of a given equilibrium quantity resummed with respect to the plasma parameter Λ should provide an alternative derivation of the high-order corrections to the most diverging graph in the $r = 0$ limit, as explained in a forthcoming work. In so doing we shall encounter truncated ocp models with the nonvanishing harmonic sum deleted for $\nu < 2$.

The present paper is organized as follows: In Sec. II, we establish the equivalence of the real-ν Coulomb interaction $r^{2-\nu}$ in configuration space with the Poisson equation solution $S_\nu k^{-2}$. We extend, to $0 \leqslant \nu \leqslant 2$, the corresponding Fourier integrals with the aid of suitably selected Tauberian factors. The high-temperature (Debye) analysis is displayed in Secs. III (pair correlation function) and IV (thermodynamic functions) up to the second-order in Λ with $\epsilon = \nu - 2$, as a running parameter. The previous ν-integer results are then re-

covered as special cases of the present extension. In Sec. V, we consider a wavepacket extension of the point-like charged particles taking into account the uncertainty principle in the short-range part of the particle—particle interaction. We conclude with the above-mentioned derivation of the particle diffusion coefficient for the two components $\nu = 2$ plasma taken in the fluid limit (Debye length $\to \infty$).

II. COULOMB INTERACTION WITH NONINTEGER DIMENSIONALITY

The compact expression (I. 3) may be immediately extended to any real ν with

$$\phi^{(\nu)}(\nu) = 1/r^{\nu-2}, \quad \nu \neq 2. \tag{II. 1}$$

Similarly, we have already obtained, in k-space, the extension (I. 7) of the Poisson solution. Therefore, it remains to introduce a noninteger Fourier relationship to connect these complementary aspects of the extended Coulomb interaction. This may be achieved through the straightforward extension of the volume integral

$$\int_0^R r^{\nu-1} \, dr \int_0^\pi \sin^{\nu-2}\phi_1 \, d\phi_1 \int_0^\pi \sin^{\nu-3} \phi_3 \, d\phi_3 \cdots$$

$$\times \int_0^\pi \sin\phi_{\nu-2} \, d\phi_{\nu-2} \int_0^\pi d\phi_{\nu-1} = \frac{2\pi^{\nu/2}}{\nu\Gamma(\nu/2)} R^\nu = \frac{S_\nu}{\nu} R^\nu \tag{II. 2}$$

and the corresponding Wilson quadratures[5]

$$(2\pi)^{-\nu} \int d^\nu k \, f(\mathbf{k} \cdot \mathbf{k}_1)$$

$$= \frac{K_{\nu-1}}{2\pi} \int_0^\infty dk \int_0^\pi d\phi \, k^{\nu-1}(\sin\phi)^{\nu-2} f(k^2, k_1 k \cos\phi), \tag{II. 3}$$

where $K_\nu = S_\nu/(2\pi)^\nu$, specialized to the generalized Fourier transforms

$$\phi^{(\nu)}(r) = \frac{K_{\nu-1}S_\nu}{(2\pi)^\epsilon} \int_0^\infty dk \, k^{\nu-1} \int_0^\pi d\phi \, \frac{(\sin\phi)^{\nu-2}}{k^2} \exp(ikr\cos\phi)$$

$$= - S_\nu S_{\nu-1}\left(\frac{2}{\nu}\right)^{\epsilon/2} \Gamma\left(\frac{1+\epsilon}{2}\right) \sqrt{\pi} \int_0^\infty dk \, k^{-1+\epsilon/2} J_{\epsilon/2}(kr)$$

$$= \frac{-1}{r^\epsilon}, \quad \epsilon > 0, \tag{II. 4}$$

and

$$\phi^{(\nu)}(k) = \frac{K_{\nu-1}}{2\pi} \int_0^\infty dr \, r^{\nu-1} \int_0^\pi d\phi \, \sin\phi^{\nu-2} \exp(-ikr\cos\phi)\phi^{(\nu)}(r)$$

$$= - 2\pi^{(1+\epsilon)/2} \left(\frac{2}{k}\right)^{\epsilon/2} \int_0^\infty dr \, r^{1-\epsilon/2} J_{\epsilon/2}(kr) \tag{II. 5}$$

$$= - \frac{S_\nu}{k^2}.$$

As in three dimensions, $\phi^{(\nu)}(k)$ is obtained through a regular quadrature for $\epsilon > 0$, while the inverse transform (II. 5) is only meaningful in the usual Tauberian limit

$$\lim_{a \to 0} \int_0^\infty dr \, e^{-ar} r^{\epsilon/2} J_{\epsilon/2}(kr)$$

$$= \frac{(2k)^{\epsilon/2}\Gamma((\epsilon+1)/2)}{(a^2+k^2)[(\epsilon+1)/2]\pi^{1/2}}\bigg|_{a=0}$$

$$\approx \frac{2^{\epsilon/2}}{k^{1+\epsilon/2}}\Gamma\left(\frac{1+\epsilon}{2}\right), \quad \epsilon > 0. \tag{II. 6}$$

1405   J. Math. Phys., Vol. 17, No. 8, August 1976

C. Deutsch   1405

We may wonder if the present analysis is able to reproduce the $\nu = 2$ logarithmic potential[1] in the $\nu = 0$ limit, and also make sense for $\epsilon < 0$, when $\phi^{(\nu)}(r)$ increases with r. These subtle points may be answered affirmatively provided the Tauberian factors are given their Debye-like analytical forms $\lambda^{\epsilon/2}K_{\epsilon/2}(\lambda r)/r^{\epsilon/2}$ and $(k^2 + \lambda^2)^{-1}$ as explained below in Sec. III. Let us resume the derivation of (II. 4) with

$$\lim_{\lambda \to 0} \int_0^\infty \frac{dk\, k^{1+\epsilon/2} J_{\epsilon/2}(kr)}{k^2 + \lambda^2}$$

$$= \lambda^{\epsilon/2} K_{\epsilon/2}(\lambda r)\Big|_{\lambda=0} \approx \frac{\Gamma(\epsilon/2)}{2} \frac{2^{\epsilon/2}}{r} \qquad \text{(II. 7)}$$

giving back the two-dimensional interaction through $K_0(r\lambda) \sim_{\lambda \to 0} \ln|\lambda r|$ provided the limit $\epsilon = 0$ is taken first. An immediate outcome of this approach is that $\lambda \sim R^{-1}$ in the thermodynamic limit, in order to reproduce the dimensionless interaction $\ln|r/R|$. R could also represent any other infinite quantity available when the plasma model is made more specific. For instance, one could also have $\lambda \sim L^{-1}$, L being the infinite length of the magnetized rods supporting the linear density of charge. This Tauberian limit makes clear that the length scaling the logarithm has to be a very large one. Apparently, this is the first unambiguous indication for that property. The relationships (II. 6), (II. 7) do confirm that the Fourier transforms (II. 4), (II. 5) may be analytically continued down to $-2 \leqslant \epsilon \lesssim 0$ when $\phi^{(\nu)}(r)$ is given the Tauberianized expression $\lim_{\lambda \to 0}(\lambda^{\epsilon/2}/r^{\epsilon/2})K_{\epsilon/2}(\lambda r)$, while $\phi^{(\nu)}(k)$ is obtained as

$$k^{-\epsilon/2} \int_0^\infty dr\, r J_{\epsilon/2}(kr) K_{\epsilon/2}(\lambda r) = (k^2 + \lambda^2)^{-1}. \qquad \text{(II. 8)}$$

III. DEBYE PAIR CORRELATION FUNCTION
A. Basic formalism and first order

Once the Fourier transform of the Coulomb potential is defined, we are allowed to develop the usual high-temperature formalism based on the perturbative analysis of the pair correlation function[1]

$$g_2(r_{12}) = \exp(\beta W_2(r_{12})), \quad \beta = (k_B T)^{-1}, \qquad \text{(III. 1)}$$

in terms of the potential of average force $w_2(r_{12})$, with respect to the dimensionless plasma parameter

$$\Lambda = \frac{\text{mean potential energy at screening distance } \lambda_D}{\text{mean kinetic energy}}$$

$$= \frac{e^2}{k_B T \lambda_D^\epsilon}. \qquad \text{(III. 2)}$$

$w_2(r_{12})$ is then explained by

$$w_2(r_{12}) = \frac{-u(r_{12})}{k_B T} + \sum_{k=1}^\infty \beta_k(r_{12})\rho^k, \qquad \text{(III. 3)}$$

in terms of the bare potential $u(r_{12}) = e^2 \phi^{(\nu)}(r_{12})$, and the simple 12-reducible cluster integrals

$$\beta_k(r_{12}) = \frac{1}{k!} \int \cdots \int d^\nu r_3 \cdots d^\nu r_{k+2} \sum^{(k)} f_{ij}, \qquad \text{(III. 4)}$$

with $\sum^{(k)}$ denoting the summation over all possible 12-irreducible cluster diagrams that can be obtained from the root points 1 and 2. k refers to the model points. $f_{ij} = \exp(-\beta u(r_{ij})) - 1$ is the usual Mayer function. Equa-

tion (III. 4) is meaningful in the $N, V \to \infty$ limit for finite k. Now, let us introduce the high-temperature approximation with the condition

$$e^2 \phi^{(\nu)}(r_{ij})/k_B T \ll 1, \quad r_{ij} \sim \rho^{-1/\nu} \qquad \text{(III. 5)}$$

without any further restriction on the number density ρ. Thus we hope to find a small parameter in terms of which the cluster expansion may be constructed with

$$u(r_{ij}) > M \gg 0, \quad r_{ij} < r_M, \qquad \text{(III. 6a)}$$

$$u(r_{ij}) \sim \epsilon', \quad r_M < r < \lambda', \qquad \text{(III. 6b)}$$

$$u(r_{ij}) \text{ decreases faster than } r^{-\nu}, \quad r > \lambda', \qquad \text{(III. 6c)}$$

and $r_M/\lambda' \ll \epsilon' \ll 1$. f_{ij} is then approximated by -1 in the region (III. 6a), of order ϵ' in (III. 6b) and negligible in (III. 6c) with $\int d^\nu r\, u(r) \sim \epsilon' \lambda'^\nu$. Now, only cases where the range of the potential is long compared to $\rho^{-1/\nu}$ ($\rho \lambda'^\nu > 1$) will be considered. Each cluster integral β_k contains k field points and l lines. The order of magnitude is given by $\epsilon'^l(\rho \lambda'^\nu)^k = \epsilon'^{l-k}(\rho \epsilon' \lambda'^\nu)^k$. Although ϵ' is by definition small, the quantity $\rho \epsilon' \lambda'^\nu$ may be large for sufficiently large λ'. It is therefore useful to regroup the cluster expansion terms for $w_2(r_{12})$ according to the value of $l - k$, the summation is over k values for fixed $l - k$. The only dimensionless parameter in the problem being $e^2 \phi^{(\nu)}(r)/k_B T$, one has to put $\epsilon' = \Lambda = (e^2/k_B T)\phi^{(\nu)}(\lambda') = e^2/k_B T \lambda_D^\epsilon$. In order to get a realistic result free from the harmonic symmetry-breaking term, we first restrict to $\nu = 2 + \epsilon$ with $\epsilon > 0$. Therefore, the first-order $(l - k = 1)$ contribution to (III. 3) is the usual Debye chain

$$\delta(r_{12}) = f(r_{12}) + \sum_{n=1}^\infty \rho^n \int \cdots \int d^\nu r_3 \cdots d^\nu r_{n+2}$$

$$\times f(r_{13}) \cdots f(r_{n+2}, 2). \qquad \text{(III. 7)}$$

The introduction of $f_{ij} \sim -\beta u(r_{ij})$ in Eq. (III. 7) leads to $[\widetilde{V}(k) = -\beta e^2 S_\nu k^{-2}]$

$$C_\nu(r_{12}) \equiv \delta(r_{12})$$

$$= (2\pi)^{-\nu} \int d^\nu k \exp(ikr\cos\theta) \frac{\widetilde{V}(k)}{1 - \rho \widetilde{V}(k)}$$

$$= \frac{K_{\nu-1}}{2\pi} \int_0^\infty dk \int_0^\pi d\theta (\sin\theta)^{\nu-2} \exp(ikr\cos\theta) \frac{\widetilde{V}(k)k^{\nu-1}}{1 - \rho \widetilde{V}(k)}$$

$$= \frac{-\Lambda}{2^{\epsilon/2}\Gamma(1 + \epsilon/2)} \frac{K_{\epsilon/2}(r/\lambda_D)}{(r/\lambda_D)^{\epsilon/2}}, \quad \lambda_D^2 = \frac{k_B T}{S_\nu \rho e^2}, \qquad \text{(III. 8)}$$

giving back at once the well-known results

$$C_3(r) = -\beta e^2 \exp(-r/\lambda_D), \quad C_2(r) = -\beta e^2 K_0(r/\lambda_D),$$

$$C_1(r) = -\beta e^2 \exp(-r/\lambda_D), \qquad \text{(III. 9)}$$

with $K_{\epsilon/2}(x)$ second kind modified Bessel of order $\epsilon/2$, and the ϵ parameter analytically continued down to $\epsilon < 0$ for the truncated ocp's with only the Coulomb repulsion retained in Eq. (I. 3). In so doing, we may also contemplate the $\nu = 0$ limit

$$C_0(r) = \frac{-2\beta e^2 \lambda_D^2}{\Gamma(0)} \frac{r}{\lambda_D} \cdot K_1\left(\frac{r}{\lambda_D}\right) = 0, \qquad \text{(III. 10)}$$

corresponding to a gas of noninteracting harmonic oscil-

lators, with the Debye approximation identical (to all orders) to the perfect gas. Equation (III.10) stresses out the topological content of the dimensionality concept already noticed in Sec. I, with the emphasis put on the local density in the neighborhood of a given point charge[7,8] rather than on the metric properties of a system, everywhere dense, and filling homogeneously the available space to it. Negative dimensionalities may also be given a meaning through first-order Debye extensions such that

$$C_{-1}(r) = -\beta e^2 \lambda_D^3 \left(\frac{2}{\pi}\right)^{1/2} \left(\frac{r}{\lambda_D}\right)^{3/2} K_{3/2}\left(\frac{r}{\lambda_D}\right) \quad \text{(III.11)}$$

B. Second order (/ -k = 2)

Once the first-order is known, the analysis may be pursued further to high orders along the same way already used[1] for $\epsilon = 0$. In so doing, we shall both emphasize the particular role played by the two-dimensional ocp, and also allow for an overwhelming simplification of the previous model calculations separately performed for integer ν. Let us first notice that the Fourier transform of the nth power of the Debye chain (III.8) is proportional to

$$\left[\int_0^\infty dr\, r^{\nu-1} \int_0^\infty d\theta\, (\sin\theta)^{\nu-2}(C_D)^n \exp(ikr\cos\theta)\right]$$

$$= \Gamma\left(\frac{\epsilon+1}{2}\right) \sqrt{\pi} \int_0^\infty dr\, r^{\nu-1} \left(\frac{2}{kr}\right)^{\epsilon/2}$$

$$\times J_{\epsilon/2}(kr)\left(\frac{-K_{\epsilon/2}(r/\lambda_D)}{2^{\epsilon/2}\Gamma(1+\epsilon/2)r^{\epsilon/2}}\right)^n, \quad \text{(III.12)}$$

a locally summable quantity in the vicinity of $r=0$ for $n < 2 + \epsilon^{-1}$; so the Mayer—Salpeter analysis outlined in Sec. IIIA may be worked out further to order $n < 2 + \epsilon^{-1}$, without embarking in the tedious Meeron resummation of the most diverging graphs[9] in the $r=0$ limit. Restricting to second-order, we have first to pay attention to the simplest 2-bubble made of two Debye lines curved between the root points 1 and 2, i.e.,

$$(2a) = \frac{\Lambda^2}{2!}\frac{K_{\epsilon/2}^2(r)}{2^\epsilon\Gamma^2(1+\epsilon/2)}, \quad r \text{ in number of } \lambda_D. \quad \text{(III.13)}$$

The next two graphs are equal to the convolution product of $(2a)$ with a single Debye line (III.8)

$$(2bc) = \frac{K_{\nu-1}}{2\pi} \int d^\nu k\, \exp(ik\mathbf{r}_{12})G(k)H(k), \quad \text{(III.14)}$$

where

$$G(k) = \frac{\Lambda^2 K_{\nu-1}}{(2\pi)^{1-\nu}2^\epsilon\Gamma(1+\epsilon/2)^2} \int d\mathbf{r}\, \exp(i k\mathbf{r})\frac{K_{\epsilon/2}^2(r)}{r^\epsilon}$$

$$= \frac{\Lambda^2 S_{\nu-1}}{2^\epsilon\Gamma(1+\epsilon/2)^2} \cdot \frac{\Gamma((1+\epsilon)/2)}{k^2} \cdot G_{2,2}^{1,2}\left(\frac{k^2}{4}\bigg|\begin{matrix}1,1+\epsilon/2\\1,1/2\end{matrix}\right), \quad \text{(III.15)}$$

given in terms of the Meijer function G_{pq}^{mn}, as detailed in Appendix A. As an example, for $\epsilon = 1$, one recovers the $\nu = 3$ bubble function

$$G(k) = \frac{4\pi\Lambda^2}{k^2} G_{2,2}^{1,2}\left(\frac{k^2}{4}\bigg|\begin{matrix}1,3/2\\1,1/2\end{matrix}\right)$$

$$= 4\pi\Lambda^2\frac{\tan^{-1}(k/2)}{2!k}, \quad \text{(III.16)}$$

obtained through

$$G_{2,2}^{1,2}\left(x\bigg|\begin{matrix}-C_1;-C_2\\a-1,b\end{matrix}\right) = \frac{\Gamma(a+C_1)\Gamma(a+C_2)}{\Gamma(a+b)}x^{a-1}$$

$$\times {}_2F_1(a+C_1, a+C_2; a+b; -x). \quad \text{(III.17)}$$

On the other hand, the Fourier transform of the Debye chain (III.8) reads

$$H(k) = \frac{-\Lambda S_{\nu-1}\Gamma((1+\epsilon)/2)\sqrt{\pi}}{\Gamma(1+\epsilon/2)}k^{-\epsilon/2} \int_0^\infty dr\, rK_{\epsilon/2}(r)J_{\epsilon/2}(kr)$$

$$= \frac{-\Lambda S_{\nu-1}\pi^{1/2}\Gamma((\epsilon+1)/2)}{\Gamma(1+\epsilon/2)(k^2+1)}, \quad \text{(III.18)}$$

giving back the $\nu = 3$ result $-4\pi(k^2+1)^{-1}$. Introducing Eqs. (III.16), (III.18) into Eq. (III.14) yields the compact expression

$$(2bc) = \frac{-\Lambda^2}{2!}\left(S_{\nu-1}\frac{\Gamma((\epsilon+1)/2)}{\Gamma(1+\epsilon/2)}\right)^3 \frac{\pi^2}{(2\pi)^\nu(2r)^{\epsilon/2}}$$

$$\times \int_0^\infty dk\frac{k^{-1+\epsilon/2}J_{\epsilon/2}(kr)}{k^2+1}$$

$$\times G_{2,2}^{1,2}\left(\frac{k^2}{4}\bigg|\begin{matrix}1,1+\epsilon/2\\1,1/2\end{matrix}\right), \quad \text{(III.19)}$$

suitable for numerical computation with r (in number of λ_D) as a running parameter. However, it appears too compact to extract in a convenient way the so important $r = 0$ and $r = \infty$ limits, and it is therefore useful to consider (III.19) under the form

$$(2bc) = \frac{-\Lambda^2}{2!}\left(S_{\nu-1}\frac{\Gamma((\epsilon+1)/2)}{\Gamma(1+\epsilon/2)}\right)^3 \frac{\pi^{3/2}}{(2\pi)^\nu r^{\epsilon/2}} \int_0^\infty \frac{dkk}{k^2+1}J_{\epsilon/2}(kr)$$

$$\times \int_0^\infty du\, u^{1-\epsilon/2}K_{3/2}^2(u)J_{\epsilon/2}(ku), \quad \text{(III.20)}$$

explained by

$$\int_0^\infty \frac{dk\, kJ_{\epsilon/2}(kr)J_{\epsilon/2}(ku)}{k^2+1} = \begin{cases} I_{\epsilon/2}(r)K_{\epsilon/2}(u), & r < u, \\ I_{\epsilon/2}(u)K_{\epsilon/2}(r), & r > u, \end{cases} \quad \text{(III.21)}$$

while the last two-legged nodal graph $(k=2)$ is a convolution of $(2a)$ with two single Debye lines located symmetrically, given as

$$(2d) = \frac{\Lambda^2}{2!}\left(\frac{S_{\nu-1}\Gamma((\epsilon+1)/2)}{\Gamma(1+\epsilon/2)}\right)^4 \frac{\pi^2}{(2\pi)^\nu r^{\epsilon/2}} \int_0^\infty \frac{dk\, kJ_{\epsilon/2}(kr)}{(k^2+1)^2}$$

$$\times \int_0^\infty du\, u^{1-\epsilon/2}K_{\epsilon/2}^2(u)J_{\epsilon/2}(ku) \quad \text{(III.22)}$$

with

$$2\int_0^\infty \frac{dk\, kJ_{\epsilon/2}(kr)J_{\epsilon/2}(ku)}{(k^2+1)^2}$$

$$= \begin{cases} uI_{\epsilon/2}(r)[uK_{\epsilon/2+1}(u) - \frac{1}{2}\epsilon K_{\epsilon/2}(u)] - K_{\epsilon/2}(u) \\ \qquad \times[\frac{1}{2}\epsilon I_{\epsilon/2}(r) + rI_{\epsilon/2+1}(r)], \quad r < u, \\ I_{\epsilon/2}(u)[rK_{\epsilon/2+1}(r) - \frac{1}{2}\epsilon K_{\epsilon/2}(r)] - K_{\epsilon/2}(r) \\ \qquad \times[\frac{1}{2}\epsilon I_{\epsilon/2+1}(u) + uI_{\epsilon/2+1}(u)], \quad r > u. \end{cases} \quad \text{(III.23)}$$

Collecting altogether Eqs. (III.13), (III.20), and (III.23),

we can write the total second-order correction to $w_2(r_{12})$ as $(\beta e^2 \lambda_D S_\nu = 1)$

$$W_2^2(r_{12}) = (2a) + 2(2be) + (2d)$$

$$= \frac{\Lambda^2}{2!\,2^\epsilon \Gamma(1+\epsilon/2)^2 r_{12}^2} \left\{ \frac{K_{\epsilon/2}^2(r_{12})}{r_{12}^{\epsilon/2}} \right.$$

$$- \frac{2\pi^{3/2}\Gamma((1+\epsilon)/2)^2 S_\nu}{\Gamma(1+\epsilon/2)} \left[\left(\int_0^{r_{12}} du\, u^{1-\epsilon/2} K_{\epsilon/2}^2(u) \right) \right.$$

$$\left. \times K_{\epsilon/2}(r_{12}) + I_{\epsilon/2}(r_{12}) \int_{r_{12}}^\infty du\, u^{1-\epsilon/2} K_{\epsilon/2}^3(u) \right]$$

$$+ \frac{S_\nu \pi^2 \Gamma((1+\epsilon)/2)^3}{2\Gamma(1+\epsilon/2)^2} \left[\int_0^{r_{12}} du\, u^{1-\epsilon/2} K_{\epsilon/2}^2(u) \right.$$

$$\times [I_{\epsilon/2}(u)(r_{12}K_{\epsilon/2+1}(r_{12}) - \tfrac{1}{2}\epsilon K_{\epsilon/2}(r_{12}))$$

$$- (\tfrac{1}{2}\epsilon I_{\epsilon/2}(u) + uI_{\epsilon/2+1}(u))K_{\epsilon/2}(r_{12})]$$

$$+ \int_{r_{12}}^\infty du\, u^{1-\epsilon/2} K_{\epsilon/2}^2(u)[-K_{\epsilon/2}(u)(\tfrac{1}{2}\epsilon I_{\epsilon/2}(r_{12})$$

$$\left. \left. + r_{12}I_{\epsilon/2+1}(r_{12})) + uI_{\epsilon/2}(r_{12})(K_{\epsilon/2+1}(u) - \tfrac{1}{2}\epsilon K_{\epsilon/2}(u))] \right] \right\}.$$

$$(III.24)$$

Equations (III. 8), (III. 24) bring into light the central role played by the $\epsilon = 2$ ocp in the high-temperature range. It is a remarkable fact that the $\nu = 0$ limit of these expressions give back the previous $\nu = 2$ results[1] worked out with $\ln|r/R|$, thus making clear that the screening process does not depend on the detailed form of $\phi^{(\nu)}(r)$, as long as the bare interaction remains long-ranged.

On the other hand, the substitutions

$$K_{1/2}(x) = (\pi/2x)^{1/2} e^{-x}, \quad I_{1/2}(x) = (2/\pi^x)^{1/2} \sinh x,$$

$$K_{3/2}(x) = (\pi/2x)^{1/2} e^{-x}(1 + 1/x),$$

$$I_{3/2}(x) = (\pi/2x)^{1/2} e^{-x}(1 + 1/x)$$

allow us to recover the $\nu = 3$ well-known correction[10]

$$W_{2,3}^2(r)/\Lambda^2 = e^{-2r}/r^2 - (1/2r)\{e^{-r}\ln 3 + e^{-r} \, \text{Ei}(-r)$$

$$- e^r \, \text{Ei}(-3r)\} + (1/8r)[(1+r)e^{-r}\ln 3$$

$$- \tfrac{4}{3}(e^{-r} - e^{-2r}) + (1+r)e^{-r} \, \text{Ei}(-r)$$

$$- (1-r)e^r \, \text{Ei}(-3r)], \quad (III.25)$$

thus illustrating the unifying power of this ν-dependent model analysis.

C. Long-range and short-range behavior

An alternative method providing a direct access to $\lim_{r\to 0} W_2(r)$ and $\lim_{r\to\infty} W_2(r)$ is afforded by the following specialization of (III. 17):

$$G_{2,2}^{1,2}\left(\frac{k^2}{4} \Big|_{1,1/2}^{1,1+\epsilon/2}\right) = \frac{\Gamma(1-\epsilon/2)}{\Gamma(3/2)} \frac{k^2}{4} \, {}_2F_1\left(1, 1-\frac{\epsilon}{2}; \frac{3}{2}; \frac{-k^2}{4}\right),$$

$$(III.26)$$

introduced in Eqs. (III. 19), (III. 22) with the result

$$(2bc, 2d) \approx \frac{\Gamma(1-\epsilon/2)}{2\pi^{1/2}r^{\epsilon/2}} \int_0^\infty \frac{dk\, k^{1+\epsilon/2}}{(k^2+1)^{1,2}} \, {}_2F_1\left(1, 1-\frac{\epsilon}{2}; \frac{3}{2}; \frac{-k^2}{4}\right).$$

$$(III.27)$$

More generally, any higher order convolution graphs with n Debye lines and p bubbles[11] may be explained as

$$\left(\frac{\Gamma(1+\epsilon/2)}{2\pi^{1/2}}\right)^p r^{-\epsilon/2} \int_0^\infty \frac{dk\, k^{1+\epsilon/2} J_{\epsilon/2}(kr)}{(k^2+1)^m} \, {}_2F_1\left(1, 1-\frac{\epsilon}{2}; \frac{3}{2}; \frac{-k^2}{4}\right)^p$$

$$\underset{r\to\infty}{\sim} \left(\frac{\Gamma(1-\epsilon/2)}{2\pi^{1/2}}\right)^p \frac{r^{m-1}}{2^{m-1}\Gamma(m)} \left(\frac{\pi}{2r}\right)^{1/2} e^{-r}. \quad (III.28)$$

The longest convolution chain with $m = n$ and $p = n - 1$ displays the characteristic ν-independent behavior

$$\left(\frac{\Gamma(1-\epsilon/2)}{2\pi^{1/2}}\right)^{n-1} \sqrt{\pi} \frac{r^{n-3/2-\epsilon/2} e^{-r}}{2^{n-3/2}\Gamma(n)}, \quad r\to\infty, \quad (III.29)$$

thus extending, to any ν value, the well-known $r^{n-2}e^{-r}$ three-dimensional asymptotic equivalent.[9,12] As a consequence, the hypernetted chain (HNC) approximation[13] to $W_2(r_{12})$ consisting of the resummation to infinity of the second-order nodal graphs, appears also valid for $\nu < 3$, as evidenced by a preliminary study of the non-convolution (Bridge) graphs which are nonnegligible, but with an increasing finite contribution, in the $r\to 0$ limit, when ν gets smaller.

The short-range behavior of $W_2(r_{12})$ is obtained from the resummation to all orders of the parallel graphs (the most important ones in the $r=0$ limit), i.e., the n-bubbles made of n Debye lines between the root points, in the exponential series

$$W_2(r_{12}) \underset{r_{12}\to 0}{\sim} -C_\nu(r_{12}) + \frac{C_\nu^2(r_{12})}{2!} + \underline{\quad} = \exp[-C_\nu(r_{12})] - 1$$

$$(III.30)$$

As a consequence, in the high-temperature range, one gets the limit behavior

$$g_2(r_{12}) \underset{r_{12}\to 0}{\sim} \exp[-C_\gamma(r_{12})] \simeq \begin{cases} \exp(-\Lambda/\epsilon \cdot 1/\epsilon), & \epsilon > 0, \\ \exp(-\Lambda/\epsilon), & \epsilon < 0, \end{cases}$$

$$(III.31)$$

while, for $\epsilon = 0$,

$$g_2(r_{12}) \underset{r_{12}\to 0}{\sim} \left(\frac{r_{12}}{\lambda_D}\right)^\Lambda.$$

The $\epsilon > 0$ result extends the already known $\gamma = 3$ short-range behavior,[14] while, for $\epsilon < 0$, the results agree with the nonrepulsive character of $\lim_{r\to 0}\phi^{(\nu)}(r)$.

IV. THERMODYNAMIC FUNCTIONS

This nodal analysis of the pair correlation function allows for a straightforward computation of the standard virial expressions for the canonical thermodynamical quantities. In contradistinction to our previous $\epsilon = 0$ studies,[1] we have to pay some attention[9] to the linearization of (III. 1) when $g_2(r_{12})$ is introduced in the virial quantities, with its first-order approximation

$$g_2(r_{12}) \sim \exp[-C_\nu(r_{12})] \simeq 1 - C_\nu(r_{12}). \quad (IV.1)$$

Such a procedure is receivable as long as the series expansion of $\exp(-C_\nu(r))$ may be term-by-term integrated with

$$\int_0^\infty dr\, r^{\nu-1} \left(\frac{K_{\epsilon/2}(r)}{r^{\epsilon/2}}\right)^P < \infty, \quad \text{all } P, \quad (IV.2)$$

1408   J. Math. Phys., Vol. 17, No. 8, August 1976

C. Deutsch   1408

a condition fulfilled for $\nu \leqslant 2$, and retaining a meaning for $\nu < 3$ when $p < 2 + \epsilon^{-1}$. Nevertheless, in this exploratory paper, we shall restrict to the first Λ-correction with $\Lambda \ll 1$, so that (IV.1) can be considered as good enough an approximation. The systematic study of the high order corrections is postponed to another work. Therefore, we obtain the virial pressure

$$\frac{P_\nu}{k_B T} = \rho - \frac{\rho^2}{2\nu k_B T}\int d^\nu \mathbf{r}\,\frac{r\partial\phi^{(\nu)}(r)}{r}g_2(r) + \frac{\rho^2}{2\nu k_B T}\int d^\nu \mathbf{r}\,r\frac{\partial}{\partial r}(r^{-\epsilon})$$

$$= \rho - \frac{\rho^2 e^2 S_\nu}{2\nu k_B T}\int_0^\infty dr\,r^\nu \frac{C_\nu(r)}{B_e^2}\frac{\partial}{\partial r}(r^{2-\nu})$$

$$= \rho\left(1 - \frac{\Lambda(1 - \epsilon/2)}{2^{1+\epsilon}\nu\Gamma(1 + \epsilon/2)}\right),\qquad \text{(IV.3)}$$

taking into account the contributions of the unscreened positive neutralizing. It easily reproduces the well-known integer equations of state

$$P_3/k_B T = \rho(1 - \Lambda/6),\quad P_2/k_B T = \rho(1 - \Lambda/4),$$

and

$$P_1 = k_B T(1 - \Lambda/2) = \rho k_B T - e\sqrt{\rho k_B T/2}\qquad\text{(IV.4)}$$

with the $\nu = 1$ result explained in a form equivalent to the two-component Prager result[22] $P_1 = 2\rho k_B T - (e/2)\times\sqrt{\rho k_B T}$ obtained from the Poisson—Boltzmann equation.

The corresponding internal energy is

$$\frac{E_\nu}{\frac{1}{2}\nu N k_B T} = 1 + \frac{\rho}{\nu k_B T}\int d^\nu \mathbf{r}\,g_2(r)\phi^{(\nu)}(r) - \frac{\rho}{\nu k_B T}\int d^\nu \mathbf{r}\,\phi^{(\nu)}(r)$$

$$= 1 - \frac{\Lambda}{2^\epsilon \nu}\frac{\Gamma(1 - \epsilon/2)}{\Gamma(1 + \epsilon/2)},\qquad\text{(IV.5)}$$

including the $\nu = 3$ result $E_3/N k_B T = \frac{3}{2}(1 - \Lambda/3)$. However, it should be noticed that our present $E_2/N k_B T = 1 - \Lambda/2$ differs from the $\ln|r/R|$ quantity[1] $E_2/N k_B T = 1 + (\Lambda/2)\times[1 - \gamma\ln(\lambda_D/2R)]$. These contrasting behaviors in the $\epsilon = 0$ limit are easily explained by the $\phi^{(\nu)}(r)$ dependence of the above virial expression. Equation (IV.2) shows that any repulsive long-range potential with the same $r\partial\phi^{(\nu)}(r)/\partial r$ yields the same equation of state. This is precisely what happens here with $\ln|r/R|$ and $1/r^\epsilon$. In contradiction, E_ν does not show up the same invariance property, thus motivating the above discrepancy. A more basic explanation for this result is afforded by the independence, already noticed in Ref. 1, of P_2 with respect to the way the thermodynamic limit ($N \to \infty$, $V \to \infty$, $\rho < -\infty$) is obtained, while E_2 strongly depends on it. The other thermodynamic quantities are easily derived from P_ν and E_ν.

The excess free energy is

$$\beta\frac{F^{exc}}{N} = \int_0^\beta d\beta'\,E(\beta') = \frac{-\Lambda\Gamma(1 - \epsilon/2)}{2^\epsilon \nu\Gamma(1 + \epsilon/2)},\qquad\text{(IV.6)}$$

while

$$\frac{C_\nu}{N} = \left(\frac{\partial E}{\partial T}\right)_N = \frac{\nu k_B}{2} + \frac{\Lambda\epsilon k_B\Gamma(1 - \epsilon/2)}{2\nu\Gamma(1 + \epsilon/2)},\qquad\text{(IV.7)}$$

and

$$\frac{S^{exc}}{N} = \frac{F^{exc} - E^{exc}}{T} = \beta^2\frac{\partial(F^{exc}/N)}{\partial\beta} = \frac{-\Lambda\epsilon\Gamma(1 - \epsilon/2)}{2^{1+\epsilon}\nu\Gamma(1 + \epsilon/2)}\qquad\text{(IV.8)}$$

with $S^{exc} = -C_\nu^{exc}$ for $\nu = 2$.

These results show that the usual polarization picture[16] of the first-order Λ approximation, with all the corrections negative except C_ν^{exc}, holds only for $\epsilon > 0$. The truncated ocp's obtained for $\epsilon < 0$ exhibit an opposite behavior which unambiguously demonstrates that the deleted harmonic terms in Eq. (I.3) are needed to secure the usual long-range polarization (screening) effect, for $\phi^{(\nu)}(r)$ increasing with r.

High order corrections to E_ν and P_ν may be obtained in a straightforward, although tedious way, by inserting the Λ corrections to $W_2(r_{12})$, obtained in Sec. III, into (IV.3) (IV.5). As far as P_ν is concerned, a more direct insight into the corresponding contributions may be obtained from its higher nodal expansion

$$\left(1 - \frac{\rho\partial}{\partial\rho}\right)\left(\rho^P\prod_{i=1}^P\int d^\nu\bar{r}_i\prod_{k=1}^P(-C_\nu(ar_{ik}\rho^{1/2})\right)$$

$$\approx \frac{\epsilon}{4(r_{12}/\lambda_D)\epsilon/2},\quad r_{12}\sim\lambda_{D'}.\qquad\text{(IV.9)}$$

Vanishing with ϵ, in terms of the first-order correction (III.8) to $g_2(r_{12})$, which indicates that the first equation of state (IV.3) is an excellent approximation for $|\epsilon| < 1$.

One may also wonder if the ring resummation performed in k-space, on the free energy itself, will reproduce the virial result (IV.6). It turns out that the Tauberian procedure involved in the resummation of the given diverging quantities is strongly ν-dependent, as shown in Appendix B, so that the expected finite result comes out only when $\nu = 3$.

High order corrections to F may also be obtained from Eq. (III.8) through

$$\beta F^{exc} = \frac{-N^2}{2V}\frac{S}{\lambda_D^\nu}\int_0^\infty dx\,x^{\nu-1}\sum_{m=3}^\infty\left(\frac{-C_{\nu(x)}}{m!}\right)^m.\qquad\text{(IV.10)}$$

V. DIFFRACTION CORRECTIONS

Apart from the dimensionality extension of the ocp model, it is also possible to consider another generalization of the Coulomb potential which amounts to replacing the pointlike $r^{-\epsilon}$ interaction by the corresponding one between spreaded out charges over quantum-mechanical wavepackets, to take into account the diffraction effects which become nonnegligible in the high-temperature regime, defined by ($\lambda = \hbar/\sqrt{m_e k_B T}$)

$$e^2/k_B T \lesssim \lambda \lesssim \lambda_{D'}.\qquad\text{(V.1)}$$

The corresponding temperature-dependent and finite at $r = 0$ effective interaction may be worked out for any ϵ, thus extending the $\nu = 3$ procedure[17] leading to the effective interaction $(e^2/r)(1 - e^{-Cr})$ with $C \sim \lambda^{-1}$, because the main steps of the corresponding k-space derivation are essentially ν-independent.

Let us consider the standard trick approximating the two-body high-temperature quantum Slater sum with the classical Gibbs expression through the ansatz[17]

$$\exp[-\beta(H_0 + H_1)] = \exp(-\beta H')\exp(-\beta H_0)G,\qquad\text{(V.2)}$$

where

$$H_0 = \sum_{i=1}^N\frac{P_i^2}{2me},\quad H' = \sum_{1\leqslant i<j\leqslant N}e^2|\bar{r}_i - \bar{r}_j|^{-\epsilon}.$$

G is thus a measure of the noncominutativity of H_0 and H' in the small β range. It is obtained as a solution of the Bloch-like equation

$$\frac{\partial G}{\partial \beta} = \exp(\beta H_0)[H_0 - \exp(\beta H')H_0\exp(-\beta H')]\exp(-\beta H_0)G.$$

(V.3)

Expanding the bracket with respect to β, one gets a series stopping exactly after the second-order, i.e.,

$$\frac{\partial G}{\partial \beta} = -\exp(\beta H_0)\left\{\beta[H', H_0] + \frac{\beta^2}{2!}[H', H_0]]\right\}\exp(-\beta H_0)G.$$

(V.4)

Restricting to the term linear in H', we get

$$G = 1 + \int_0^\beta \frac{d}{d\beta_1}[\exp(\beta_1 H_0)H'\exp(-\beta_1 H_0)]d\beta_1,$$

(V.5)

which allows the density operator $\rho = \exp[\beta(F - H)]$, with $H = H_0 + H'$, to be written as

$$\rho\exp(-\beta F) = \exp(-\beta H)$$
$$= \exp(-\beta H')\exp(-\beta H_0) + \exp(-\beta H')\int_0^\beta \beta_1\frac{d}{d\beta_1}$$
$$\times\{\exp[(\beta_1 - \beta)H_0]H'\exp[-(\beta_1 - \beta)H_0]\}$$
$$\times\exp(-\beta H_0)\,d\beta_1,$$

(V.6)

simplified through $\beta' = \beta_1 - \beta$ with

$$\exp(\beta' H_0)H'\exp(-\beta' H_0)$$
$$= \frac{1}{2}\sum_{i\neq j}\exp(\beta' H_0)(2\pi)^{-\nu}\int d^\nu k\,\exp(i\mathbf{k}\cdot\mathbf{r}_{ij})\phi^{(\nu)}(k)\exp(-\beta' H_0)$$

(V.7)

and

$$F_{ij} = \exp(\beta' H_0)\exp[i\mathbf{k}\cdot(\mathbf{r}_i - \mathbf{r}_j)]\exp(-\beta' H_0)$$
$$= \exp[i\mathbf{k}\cdot(\mathbf{r}_i - \mathbf{r}_j)]\exp(\beta'\hbar\mathbf{k}/m_e\cdot(\mathbf{P}_i - \mathbf{P}_j)]$$
$$\times\exp(\beta'\hbar^2 k^2/m_e),$$

(V.8)

explained in Appendix C, in the form

$$\exp(-\beta H) = \exp(-\beta H')\exp(-\beta H_0) + \frac{1}{2(2\pi)^\nu}\sum_{i\neq j}\exp(-\beta H')$$
$$\times\int_0^\beta\int_{\mathbf{k}}\{\phi^{(\nu)}(\mathbf{k})\exp(i\mathbf{k}\cdot\mathbf{r}_{ij}$$
$$\times\beta_1\frac{d}{d\beta_1}\left[\exp[(\beta_1 - \beta)\hbar\mathbf{k}/m_e\cdot(\mathbf{P}_i - \mathbf{P}_j)]\right.$$
$$\times\exp[(\beta_1 - \beta)\hbar^2 k^2/m_e]\bigg]\exp(-\beta H_0)\}\,d\beta_1\,d\mathbf{k}. \quad(V.9)$$

The corresponding canonical partition function $Z = \mathrm{Tr}\exp(-\beta H)$ is explained by

$$\langle\bar{r}_1\cdots\bar{r}_N|\exp\left(-\tfrac{1}{2}\beta\right)\sum_{i=1}^N P_i^2\right)|\bar{r}_i\cdots\bar{r}_N\rangle$$
$$= [(S_\nu/2)\Gamma(1 + \epsilon/2)(2m_e/\beta\hbar^2)^{1+\epsilon/2}]^N$$

(V.10)

and also

$$\langle\bar{r}_i\cdots\bar{r}_N|\exp\left(\frac{\beta'\hbar\mathbf{k}}{m_e}\cdot(\mathbf{P}_i - \mathbf{P}_j)\right)\exp\left(-\frac{\beta}{2}\sum_{i=1}^N\frac{P_i^2}{m_e}\right)|\bar{r}_i\cdots\bar{r}_N\rangle$$
$$= \left(\frac{S_\nu}{2}\Gamma\left(1 + \frac{\epsilon}{2}\right)\left(\frac{2m_e}{\beta\hbar^2}\right)^{1+\epsilon/2}\right)^N\exp(\hbar^2\beta'^2 k^2/\beta m_e) \quad(V.11)$$

so that

$$\langle\bar{r}_i\cdots\bar{r}_N|\exp(-\beta H)|\bar{r}_i\cdots\bar{r}_N\rangle$$
$$= \left[\frac{S_\nu}{2}\left(\frac{2m_e}{\beta\hbar^2}\right)^{1+\epsilon/2}\Gamma\left(1 + \frac{\epsilon}{2}\right)\right]^N\exp(-\beta/2)\sum_{i\neq j}\Phi_{(r_{ij})}^{(\nu)}$$
$$\times\left\{1 + \frac{1}{2(2\pi)^\nu}\cdot\sum_{i\neq j}\int_0^\beta\beta_1\frac{d}{d\beta_1}\int_{\mathbf{k}}\Phi^{(\nu)}(k)\exp(i\mathbf{k}\cdot\mathbf{r}_{ij})\right.$$
$$\times\exp\left(\frac{\hbar^2 k^2}{m_e}(\beta' + \beta'^2/\beta)\right)d\mathbf{k}\,d\beta_1 + \text{---}\bigg\} \quad(V.12)$$

Now, let us simplify the β_1-quadrature with the new variable $\alpha = \beta'/\beta$ and $\beta' + \beta'^2/\beta = \beta\alpha(\alpha - 1)$, and approximate the high order quantum corrections through the exponential expansion

$$\langle\bar{r}_i\cdots\bar{r}_N|\exp(-\beta H)|\bar{r}_i\cdots\bar{r}_N\rangle$$
$$= \left[\frac{S_\nu}{2}\left(\frac{2m_e}{\beta\hbar^2}\right)^{1+\epsilon/2}\Gamma\left(1 + \frac{\epsilon}{2}\right)\right]^N\exp\bigg\{-\frac{\beta}{2(2\pi)^\nu}$$
$$\times\sum_{i\neq j}\int_{\mathbf{k}}\Phi^{(\nu)}(k) - \int_0^1\alpha\frac{d}{d\alpha}\Phi^{(\nu)}(k)$$
$$\times\exp\left(\frac{-\beta\hbar^2 k^2\alpha(1 - \alpha)}{m_e}\right)\exp(i\mathbf{k}\cdot\mathbf{r}_{ij})\,d\mathbf{k}\,d\alpha\bigg\}$$
$$= \left(\frac{S_\nu}{2}\left(\frac{2m_e}{\beta\hbar^2}\right)^{1+\epsilon/2}\Gamma\left(1 + \frac{\epsilon}{2}\right)\right)^N\exp\left(\frac{-\beta}{2}\sum_{i\neq j}W(r_{ij})\right),$$

(V.13)

obtained after an integration by parts with respect to α, and explained with the temperature-dependent effective interaction

$$W(r_{ij}) = (2\pi)^{-\nu}\int_{\mathbf{k}}\Phi^{(\nu)}(k)$$
$$\times\int_0^1\exp\left(\frac{-\beta\hbar^2 k^2\alpha(1 - \alpha)}{m_e}\right)d\alpha\exp(i\mathbf{k}\cdot(\mathbf{r}_i - \mathbf{r}_j)d\mathbf{k},$$

(V.14)

which has the Fourier-transform

$$W(k) = \Phi^{(\nu)}(k)\int_0^1 d\alpha\exp\left(\frac{-\hbar^2 k^2\alpha(1 - \alpha)\beta}{m_e}\right)$$
$$= \Phi^{(\nu)}(k)\int_0^1 d\alpha\exp(-\lambda^2 k^2\alpha(1 - \alpha))$$
$$= \Phi^{(\nu)}(k)_1 F_1(\tfrac{1}{2}, \tfrac{3}{2}, -\tfrac{1}{4}\lambda^2 k^2). \quad(V.15)$$

Equations (V.14), (V.15) extend to any ν, the $\nu = 3$ effective interaction finite at $r = 0$, in the form

$$W(r) = \left(\frac{2}{r}\right)^{\epsilon/2}\Gamma\left(\frac{\epsilon + 1}{2}\right)\pi^{1/2}\int_0^\infty dk\,k^{-1+\epsilon/2}{}_1F_1(1; \tfrac{3}{2}; -\tfrac{1}{4}\lambda^2 k^2)$$
$$\times J_{\epsilon/2}(kr) = \Gamma\left(\frac{\epsilon + 1}{2}\right)\left(\frac{2}{r}\right)^{\epsilon/2}\frac{\pi}{4}\cdot G_{2,3}^{2,1}\left(\frac{r^2}{\lambda^2}\bigg|\begin{matrix}1,3/2\\ \epsilon/4,1,0\end{matrix}\right),$$
$$\epsilon > 0, \quad(V.16)$$

including the $\epsilon = 1$ result[17]

$$W_3(r) = r^{-1}(1 - \exp(-r^2/\lambda^2)) + (\pi^{1/2}/\lambda)\,\mathrm{Erf}(r/\lambda)$$
$$\approx r^{-1}(1 - e^{-Cr}) \quad(V.17)$$

with $\mathrm{Erf}(x) = 1 - 2/\sqrt{\pi}\int_0^x\exp(-t^2)\,dt$ and $C \approx \lambda^{-1}$.

The Tauberian generalization of the Fourier recipro-

city relationship (II.4), (II.5) allow to extend the effective interaction (V.16) down to $\epsilon \leqslant 0$ with

$$W(r) = \lim_{\lambda' \to 0} \left(\frac{2}{r}\right)^{\epsilon/2} \Gamma\left(\frac{\epsilon+1}{2}\right) \pi^{1/2} \int_0^\infty \frac{dk\, k^{1+\epsilon/2}}{k^2 + \lambda'^2}$$

$$\times {}_1F_1(1; \tfrac{3}{2}; -\tfrac{1}{4}\lambda^2 k^2) J_{\epsilon/2}(kr). \qquad (V.18)$$

VI. PARTICLE TRANSPORT IN A STRONGLY MAGNETIZED TWO-COMPONENT $\nu = 2$ PLASMA

As a first application to current plasma physics, we consider the evaluation of the time-independent transport coefficients such as the particle diffusion coefficient on the corresponding d.c. conductivity in a two-component $\nu = 2$ symmetrical plasma taken in the fluid limit, i.e., with close collisions and any other discrete processes neglected. The given analysis is performed in momentum space with $k_{\min} \leqslant k \leqslant \lambda_D^{-1}$, and $k_{\min} \sim R^{-1}$. We suppose the system to be confined by a very strong magnetic field B, which enables us to modelize and restrict very efficiently the time dependence of the particles trajectories. The $\nu = 2$ version of that model has already received a considerable attention,[3,18-20] because it is expected to provide a realistic approximation of the strongly magnetized $\nu = 3$ plasma with the particles dynamics restricted perpendicularly to B. For instance, the transverse velocity diffusion coefficient in the $\nu = 2$ model of charged filaments[3,18] aligned along B is given as (c = speed of light)

$$D = \frac{c^2}{B^2} \int_0^\infty \langle \mathbf{E}(0) \cdot \mathbf{E}(\tau) \rangle \, d\tau \qquad (VI.1)$$

in terms of the autocorrelation function of the electric field $\mathbf{E}(\tau)$ seen by a test charge at time τ, with the guiding center approximation

$$X(t) = c \int_0^t \frac{\mathbf{E}(\tau) \Lambda \mathbf{B}}{B^2} \, d\tau.$$

The bracket in Eq. (VI.1) is explained as[3,18]

$$\langle \mathbf{E}(0) \cdot \mathbf{E}(\tau) \rangle$$

$$= \sum_{\mathbf{k}_1, \mathbf{k}_2} \langle \mathbf{E}_{\mathbf{k}_1}(0) \cdot \mathbf{E}_{\mathbf{k}_1}(\tau) \exp(i\mathbf{k}_1 \cdot \mathbf{X}(\tau)) \rangle \qquad (VI.2)$$

where

$$\mathbf{E}(\tau) = \sum_{\mathbf{k}} \mathbf{E}_{\mathbf{k}}(\tau) \exp(i\mathbf{k} \cdot \mathbf{x}(\tau)). \qquad (VI.3)$$

To go farther, one neglects the correlation between the positions of the test particle and those of the background plasma [the $\mathbf{X}(\tau)$]. That assumption reduces (VI.2) to

$$\langle \mathbf{E}(0) \cdot \mathbf{E}(\tau) \rangle \simeq \sum_{\mathbf{k}_1, \mathbf{k}_2} \langle \mathbf{E}_{\mathbf{k}_1}(0) \cdot \mathbf{E}_{\mathbf{k}_2}(\tau) \rangle$$

$$\times \langle \exp(i\mathbf{k} \cdot \mathbf{X}(\tau)) \rangle \qquad (VI.3')$$

for a spatially uniform system with $\mathbf{E}_{-\mathbf{k}}(t) = \mathbf{E}_{\mathbf{k}}(t)$, since E is real. Then the use of the central limit theorem allows us to explain (VI.3) as

$$\langle \mathbf{E}(0) \cdot \mathbf{E}(\tau) \rangle$$

$$= \sum_{\mathbf{k}} \langle |\mathbf{E}_{\mathbf{k}}|^2 \rangle \exp\left\{-\frac{c^2 k^2}{2B^2} \int_0^\tau d\tau_1 \int_0^\tau d\tau_2 \langle \mathbf{E}(\tau_1) \mathbf{E}(\tau_2) \rangle \right\},$$

$$\qquad (VI.4)$$

where $\langle |\mathbf{E}_k|^2 \rangle$ is now a stationary quantity obtainable from the ocp equilibrium properties. Let us define[3,18]

$$R(\tau) = \frac{c^2}{2B^2} \int_0^\tau d\tau_1 \int_0^\tau d\tau_2 Q(\tau_2 - \tau_1),$$

$$Q(\tau) \equiv 2^{-1} \langle \mathbf{E}(0) \cdot \mathbf{E}(\tau) \rangle$$

so that Eq. (VI.4) may be expressed in the form

$$\frac{2B^2}{C^2} \frac{d^2R}{d\tau^2} = \sum_{\mathbf{k}} \langle |\mathbf{E}_{\mathbf{k}}|^2 \rangle \exp(-2k^2 R(\tau)), \qquad (VI.5)$$

which has the first integral $(dR(0)/d\tau = 0)$

$$\frac{1}{2}\left(\frac{dR}{d\tau}\right)^2 - \frac{c^2}{2B^2} \sum_{\mathbf{k}} \langle |E_{\mathbf{k}}|^2 \rangle \frac{1 - \exp(-2k^2 R(\tau))}{2k^2} = 0. \qquad (VI.6)$$

We can extract the long time behavior analytically, by noting that $R(\tau) \to \infty$ as $\tau \to \infty$, so that

$$\left(\frac{dR}{d\tau}\right)^2 - \frac{c^2}{B^2} \frac{\langle |\mathbf{E}_{\mathbf{k}}|^2 \rangle}{2k^2}, \quad \tau \to \infty, \qquad (VI.7)$$

or

$$D(t) = \frac{2c^2}{B^2} \int_0^t Q(\tau) \, d\tau = \frac{2dR(t)}{dt}$$

approaches the limit

$$D(\infty) = 2\frac{c^2}{B^2} \left[\sum_{\mathbf{k}} \frac{\langle |E_{\mathbf{k}}|^2 \rangle}{2k^2}\right]^{1/2} \qquad (VI.8)$$

with the characteristic B^{-1}-(Bohm) dependence. We may estimate Eq. (VI.8) for a large volume $V \approx R^2$ by replacing the sum over discrete values of k by an integral over continuous k according to the prescription

$$\sum_{\mathbf{k}} \langle |\mathbf{E}_{\mathbf{k}}|^2 \rangle - \int d\mathbf{k} \langle |\mathbf{E}_{\mathbf{k}}|^2 \rangle \qquad (VI.9)$$

using the thermal spectrum[19]

$$\langle |\mathbf{E}_{\mathbf{k}}|^2 \rangle = k_B T / \pi ((k\lambda_D)^2 + 1). \qquad (VI.10)$$

Equation (VI.8) becomes

$$D(\infty) = 2\left[\frac{c^2}{B^2} \frac{k_B T}{\pi} \int \frac{d\mathbf{k}}{2k^2} \frac{1}{1 + k^2\lambda_D^2}\right]^{1/2}. \qquad (VI.11)$$

According to the previous discussion given in Sec. II, concerning the Tauberian extension of the relations (II.4), (II.5), it appears now convenient to replace the bare Coulomb interaction k^{-2} in Eq. (VI.11) with $\lim_{\lambda \to 0}(k^2 + \lambda^2)$, so that

$$D \equiv D(\infty) = \lim_{\lambda \to 0} \frac{c}{B} \left[\int_{k_{\min}}^{k_{\max}} \frac{dk\, k}{k^2 + \lambda^2} \cdot \frac{1}{k^2\lambda_D^2 + 1}\right]^{1/2} \qquad (VI.12)$$

with the usual fluid-limit cutoffs $k_{\min} = 2\pi/R$ and $k_{\max} = \lambda_D^{-1}$. Therefore, the transport properties of the real two-component system may correspond to the symmetrical superposition of two ocp's (electron + ion). If we were to take the $\lambda \to 0$ limit in a crude way, we would merely reproduce the already known diverging result[3,18]

$$D = \sqrt{2} \frac{ck_B T}{eB} \left[\Lambda \ln\left(\frac{R}{2\pi\lambda_D}\right)\right]^{1/2} \qquad (VI.13)$$

So, we invoke previous two-dimensional numerical analysis[20] showing clearly that the $\nu = 2$ transport coefficients are more conveniently shielded by the average diameter R of the system, rather than by λ_D, and the Tauberian

parameter λ may be taken $\sim \lambda_D^{-1}$ with $\lambda_D \to \infty$ $(T \to \infty)$ and $\lambda_D \ll R$ in the fluid limit with

$$\lim_{\lambda \to 0, |\lambda - \lambda_D^{-1}| \to 0} \int \frac{dk\, k}{(k^2 + \lambda^2)(k^2 \lambda_D^2 + 1)}$$

$$= \frac{1}{2\lambda_D^2}\left[\frac{-\partial}{\partial \lambda^2} \ln(k_{max}^2 + \lambda^2)_{\lambda^2 = \lambda_D^{-2}} + \frac{\partial}{\partial \lambda^2}\ln(k_{min}^2 + \lambda^2)_{\lambda^2 = \lambda_D^{-2}} \right]$$

$$= \frac{1}{2}\left(-\frac{1}{2} + \frac{1}{k_{min}^2 \lambda_D^2 + 1}\right), \qquad (VI.14)$$

Introducing (VI.14) into Eq. (VI.12) yields

$$D = \sqrt{2}\frac{c}{eB} k_B T \Lambda^{1/2}\left[-\frac{1}{2} + \frac{1}{k^2 \lambda_D^2 + 1}\right]^{1/2}, \quad k_{min}\lambda_D \ll 1$$

$$\approx \frac{c}{eB} k_B T \Lambda^{1/2} \qquad (VI.15)$$

the usual Bohm-like result without the $\ln V$ divergence. The same result holds for the d.c. conductivity σ, a result expected from the Einstein propertionality between D and σ, valid when the recombination processes may be neglected.

As a consequence, we can write the previous diverging result[20] in the form

$$\sigma \approx \frac{c\rho e}{B} \Lambda^{1/2}\left(-\frac{1}{2} + \frac{1}{k_{min}^2 \lambda_D^2 + 1}\right)^{1/2}. \qquad (VI.16)$$

On the other hand, it is easy to convince oneself that the potential arising from a moving test charge[20] cannot be screened that way in the fluid limit.

APPENDIX A

Here, we explain the radial integral involved in Eq. (III.15)

$$I = \int_0^\infty dr\, r^{1-\epsilon/2} K_{\epsilon/2}^2(r) J_{\epsilon/2}(kr) \qquad (A1)$$

with the aid of the McDonald relationship

$$2K_\nu^2(r) = \int_0^\infty \frac{dv}{v}\exp\left(\frac{-V - r^2}{2\ V}\right)K_\nu\left(\frac{r^2}{V}\right) \qquad (A2)$$

in the form

$$I = 2^{-1}\int_0^\infty \frac{dv}{v}\exp(-V/2)\int_0^\infty dr\, r^{1-\epsilon/2}$$

$$\times \exp(-r^2/V)J_{\epsilon/2}(kr)K_{\epsilon/2}\left(\frac{r^2}{v}\right) \qquad (A3)$$

evaluated through

$$\int_0^\infty dr\, r^{1-\epsilon/2} J_{\epsilon/2}(kr)K_{\epsilon/2}\left(\frac{r^2}{v}\right)\exp(-r^2/V)$$

$$= \left(\frac{\pi}{2}\right)^{1/2} 2^{(3-\epsilon)/2}k^{-2+\epsilon/2}\, G_{1\ 2}^{1\ 1}\left(\frac{k^2 v}{8}\bigg|\begin{matrix}1+\epsilon/2\\1,1/2\end{matrix}\right) \qquad (A4)$$

and

$$\int_0^\infty \frac{dv}{v}\exp(-v/2)G_{12}^{11}\left(\frac{k^2 v}{8}\bigg|\begin{matrix}1+\epsilon/2\\1,1/2\end{matrix}\right) = G_{2,2}^{1,2}\left(\frac{k^2}{4}\bigg|\begin{matrix}1,1+\epsilon/2\\1,1/2\end{matrix}\right) \qquad (A5)$$

with

$$I = \frac{\sqrt{\pi}}{2^{\epsilon/2}}k^{-2+\epsilon/2}G_{2,2}^{1,2}\left(\frac{k^2}{4}\bigg|\begin{matrix}1,1+\epsilon/2\\1,1/2\end{matrix}\right),$$

r in number of λ_D. \hfill (A6)

APPENDIX B: RING RESUMMATION FOR F exc

It is also possible to perform directly the long-range resummation of the Coulomb potential, within the frame of the so-called ring resummation method,[15,21] on the cluster integrals building up directly the excess free energy. $\phi^{(\nu)}(r)$ indicates that the most divergent cluster integrals involving f particles are those of the type $f_{12}f_{23}\cdots f_{j-1,j}f_{j1}$; each particle is joined to just two others, the whole cluster forming a ring. Furthermore, when the f functions are expanded in powers of $\phi^{(\nu)}(r)$, the first term is always the most divergent, when the relative distance $r_{ij} \to \infty$. The virial coefficient B_j is thus approximated by the expression[15,16]

$$B_j = \frac{(-)^j(j-1)!}{2}\cdot\frac{\beta^j}{j!\ V}\int\cdots\int d^\nu r_1\cdots d^\nu r_j \phi^{(\nu)}(r_{12})$$

$$\times \phi^{(\nu)}(r_{23})\cdots\phi^{(\nu)}(r_{ji}), \qquad (B1)$$

where the combinatorial factor $(j-1)!/2$ is the number of ways the j particles can be placed on the ring.

This convolution integral can be treated with the aid of the reciprocity relations (II.5), (II.6) together with

$$\delta^{(\nu)}(k) = (2\pi)^\nu\int d^\nu r\, \exp(i\mathbf{k}\cdot\mathbf{r})$$

to yield

$$B_j \underset{\lambda \to 0}{\approx} \frac{(-)^j S_\nu(S_\nu\beta e^2)^j}{2j(2\pi)^\nu}\int_0^\infty \frac{dk\, k^{1+\epsilon}}{(k^2 + \lambda^2)^j} \qquad (B2)$$

where λ is the Tauberian parameter. To eliminate the long-range $(k=0)$ divergence, we sum over j before integrating over k, so that we retain the ring contribution to the free energy

$$B\,F_{ring} = \sum_{j=2}^\infty B_j\rho^j$$

$$= \frac{S_\nu}{2(2\pi)^\nu}\int_0^\infty dk\, k^{\nu-1}\sum_{j=2}^\infty j^{-1}\left(\frac{-S_\nu\beta e^2\rho}{k^2 + \lambda^2}\right)^j$$

$$= \frac{-S_\nu N G(\eta)}{2(2\pi)^\nu\rho\lambda_D^\nu} \qquad (B3)$$

where[15] $(\eta = \lambda\,\lambda_D)$

$$G(\eta) = \sum_{n=2}^\infty \frac{(-)^n}{n}\frac{dx\, x^{\nu-1}}{(x^2 + \eta^2)^n}$$

$$= \frac{1}{2i\pi}\int_{-i\infty}^{i\infty}\frac{ds\, \pi}{s\,\sin\pi s}\int_0^\infty \frac{dx\, x^{\nu-1}}{(x^2 + \eta^2)^s}, \quad 1 < \mathrm{Re}\,s < 2, \qquad (B4)$$

using the Mellin transform representation of the discrete sum. With the change of variable $x = \eta\tan\theta$, the integration becomes

$$\eta^{\nu-2s}\int_0^\pi d\theta\,\sin^2\theta\cos\theta^{2s-\nu-1} = \frac{\eta^{\nu-2s}}{2}\frac{\Gamma(\frac{3}{2})\Gamma(s-1-\epsilon/2)}{\Gamma(s+\frac{1}{2}-\epsilon/2)} \qquad (B5)$$

so that we have

$$G(\eta) = \frac{\Gamma(\frac{3}{2})}{4i\pi}\int_C\frac{ds\,\pi\eta^{-2s}}{s\,\sin\pi s}\cdot\frac{\Gamma(s-1-\epsilon/2)}{\Gamma(s+\frac{1}{2}-\epsilon/2)} \qquad (B6)$$

to generate a series in power of η appropriate to $\eta \ll 1$. The contour defined in (B6) is closed to the left. Thus, the contour C encloses the entire axis to the left of and

including the points $s = 1 + \epsilon/2$. Similarly, to generate a series in η^{-1} appropriate for large η, the contour is to the right. The integrand of (B6) has simple poles at $s = 1 + \epsilon/2, \frac{1}{2} + \epsilon/2, \ldots, -m+1+\epsilon/2$, from $\Gamma(s - 1 - \epsilon/2)$ and two additional simple poles at $s = 1$ and 0 from $(\sin\pi s)^{-1}$ when $\epsilon = 1$, or a simple pole at $s = 1$ and a double pole at $s = 0$ for $\epsilon < 1$. Thus we see the particular analytic structure of the ring sum for $\nu = 3$.

Let us first consider the general case $\epsilon < 1$

$$G(\eta) = \frac{\Gamma(\frac{3}{2})\pi}{2} \left\{ \frac{-(1+\eta^2)^{\nu/2}}{\nu \sin\Gamma(1+\epsilon/2)} - 2\ln\eta \, \eta^\nu \frac{\Gamma(1-\epsilon/2)}{\Gamma((1-\epsilon)/2)} \right.$$

$$+ \eta^\nu \left[\frac{\Psi(-1-\epsilon/2)\Gamma(-1-\epsilon/2) - \Psi((1-\epsilon)/2)\Gamma(1-\epsilon/2)}{\pi\Gamma(1-\epsilon/2)} \right]$$

$$+ \left. \frac{\eta^\nu}{\pi s} \frac{\Gamma(s-1-\epsilon/2)}{\Gamma(s+1-\epsilon/2)} \right|_{s=0} \frac{-\eta^{\nu-2}}{\pi} \frac{\Gamma(-1-\epsilon/2)}{\Gamma((1-\epsilon)/2)} \right\} . \qquad (B7)$$

Equation (B7) diverges if the Tauberian limit $\eta = 0$ is taken (as usual) at the end. For $\nu = 3$, the $\epsilon = 0$ pole becomes simple, and the above expression is replaced by the much simpler one

$$G(\eta) = (\pi/3)(1+\eta^2)^{3/2} - 3\eta/2 - \eta^3 \qquad (B8)$$

with $G(0) = \pi/3$ and $\beta F_{\text{ring}} = -N\Lambda/3$, in accord with the virial result (IV.6). This study makes clear that the usual ring resummation procedure is quite a touchy procedure, which cannot be extended without appropriate modification to any $\nu \neq 3$.

APPENDIX C

The relationship (V.8) is easily derived from the usual quantum-mechanical expansion

$$\exp\left(\frac{\beta'}{2m_e} P_i^2\right) \exp(itq_i) \exp\left(-\frac{\beta'}{2m_e} P_i^2\right)$$

$$= \exp(itq_i) + \frac{\beta'}{2m_e} [P_i^2, \exp(itq_i)]$$

$$+ \frac{1}{2!}\left(\frac{\beta'}{2m_e}\right)^2 [P_i^2, [P_i^2, \exp(itq_i)]] + \frac{}{} \qquad (C1)$$

for the canonical variables (q_i, p_i) of the electron i.

If f depends only on q_i, one respectively obtains

$$P_i^2 f - f P_i^2 = -\hbar^2 \frac{\partial^2 f}{\partial q_i^2} - 2i\hbar \frac{\partial f}{\partial q_i} P_i \qquad (C2)$$

and $(f = \exp(itq_i))$

$$P_i^2 \exp(itq_i) - \exp(itq_i)P_i^2$$

$$= \hbar^2 k^2 \exp(itq_i) + 2\hbar k \exp(itq_i)P_i \qquad (C3)$$

allowing us through

$$[P_i^2, [P_i^2, \exp(itq_i)]] = \exp(itq_i)(\hbar^2 k^2 + 2\hbar k P_i)^2 \qquad (C4)$$

to represent the rhs of (C.1) with an exponential expansion, so that

$$\exp\left(\frac{\beta'}{2m_e} P_i^2\right) \exp(itq_i) \exp\left(\frac{-\beta'}{2m_e} P_i^2\right)$$

$$= \exp(itq_i) \exp\left[\frac{\beta'}{m_e}\left(\hbar k P_i + \frac{\hbar^2 k^2}{2}\right)\right] \qquad (C5)$$

yields the rhs of Eq. (V.8), when the reduced mass $m_e/2$ is introduced in it.

[1] C. Deutsch and M. Lavaud, Phys. Rev. A 9, 2598 (1974); E.H. Hauge and P.C. Hemmer, Physica Norvegica 5, 209 (1971); R.R. Sari, D. Merlini, and R. Calinon, EPL-preprint, 1975.

[2] H. Kunz, Ann. Phys. (N.Y.) 85, 303 (1974); R.J. Baxter, Phys. Fluids 1, 38 (1964).

[3] D. Montgomery, Proceedings Les Houches Summer School of Theoretical Physics, July 1972 (Gordon and Breach, New York, 1975).

[4] D.R. Herrick and F.H. Stillinger, Phys. Rev. A 11, 42 (1975).

[5] K. Wilson, Phys. Rev. Lett. 28, 548 (1972).

[6] V. De Alfaro and T. Regge, Potential Scattering (North-Holland, Amsterdam, 1965).

[7] H. Poincaré, Rev. Metaphys. Morale (Paris) 20, 486 (1912).

[8] W. Hurewicz and H. Wallman, Dimension Theory (Princeton U.P., Princeton, N.J., 1950).

[9] E.G.D. Cohen and T.J. Murphy, Phys. Fluids 12, 1404 (1969).

[10] H.E. DeWitt, Phys. Rev. 140, A466 (1965).

[11] Actually we mean 2-bubble.

[12] F. Del Rio and H.E. DeWitt, Phys. Fluids 12, 791 (1969); D.J. Mitchell and B.W. Ninham, Phys. Rev. 176, A466 (1968).

[13] C. Deutsch and Y. Furutani, J. Phys. A 8, L83 (1975); M.S. Cooper, Phys. Rev. A 7, 1 (1973).

[14] M.S. Cooper and H.E. DeWitt, Phys. Lett. A 40, 391 (1972); C. Deutsch, Phys. Lett. A 53, 183 (1975).

[15] S.G. Brush, H.E. DeWitt, and J.G. Trulio, Nucl. Fusion 3, 5 (1963).

[16] R. Balescu, Statistical Mechanics of Charged Particles (Interscience, Wileys Ed., New York, 1963), Chap. 12.

[17] G. Kelbg, S. Bose 70th Birthday Commemoration Volume (Calcutta, 1965), p. 100.

[18] J.B. Taylor and B. McNamara, Phys. Fluids 14, 1492 (1971).

[19] J.B. Taylor, Phys. Rev. Lett. 32, 199 (1974).

[20] D. Montgomery and F. Tappert, Phys. Fluids 15, 683 (1972).

[21] R. Brout, Phys. Rev. 115, 824 (1959).

[22] S. Prager, Adv. Chem. Phys. 4, 201 (1963).

Stability and instability conditions for nonlinear evolutional equations in Hilbert spaces

P. K. C. Wang

School of Engineering and Applied Science, University of California, Los Angeles, California 90024
(Received 24 February 1976)

Sufficient conditions for stability, global asymptotic stability, and explosive instability are established for a class of nonlinear evolutional equations defined in Hilbert spaces by using certain relations between an abstract function and its Gateaux differential. These results are applied to specific forms of nonlinear evolutional equations arising from physics, in particular, a finite-dimensional system of complex ordinary differential equations, functional differential equations, and systems of complex partial differential equations describing nonlinear diffusion or wave phenomena.

I. INTRODUCTION

The equations governing the dynamics of many physical systems such as finite systems of nonlinearly interacting particles, and continuum systems involving nonlinear wave propagation, diffusion and other phenomena can be formulated as evolutional equations in suitable abstract vector spaces. This formulation is particularly useful when one is interested in the stability or the qualitative behavior of the solutions. Recently, attempts have been made in developing general stability theories for abstract nonlinear dynamical systems.[1-3] Also, a considerable amount of attention has been focused on seeking exact explicit solutions of specific classes of nonlinear partial differential equations arising from physics.[4,5] In general, the solutions to a nonlinear evolutional equation may not be well-behaved at all time. Explosive instabilities or finite escape-time phenomena may occur in which some of the solutions become unbounded in finite time.[6-8] In this paper, we develop sufficient conditions for stability or instability of solutions for a class of nonlinear evolutional equations defined in Hilbert spaces. These conditions are expressed in terms of the Gateaux differentials of the equations' right-hand sides. They may be used to determine the existence or nonexistence of explosive instabilities. The applications of the main results are illustrated by specific examples.

II. PRELIMINARIES

Let $I = [t_0, \infty[$ and H be a complex Hilbert space with inner product $(\circ, \cdot)$ and its induced norm $\|\circ\|$. We consider the initial-value problem for the following nonlinear evolutional equation:

$$d\mathbf{u}/dt = \mathbf{f}(t, \mathbf{u}), \quad t > t_0, \tag{1}$$

$$\mathbf{u}(t_0) = \mathbf{u}_0, \tag{2}$$

where $\mathbf{f}$ is a given function from $I \times U$ into H with U being a linear subspace of H, and $\mathbf{u}_0$ is a given element in U. A function $\mathbf{u}: I \to H$ is called a solution of (1) with initial data $\mathbf{u}_0$ at $t = t_0$, if $\mathbf{u} \in C_1(I;H)$ and $\mathbf{u}(t) \in U$ for each $t \in I$ with $\mathbf{u}(t_0) = \mathbf{u}_0$. [$C_m(I;H)$ is the Banach space of all m-times continuously differentiable H-valued functions on I with the norm given by $\|\|\mathbf{u}\|\| = \sum_{i=0}^{m} \sup_{t \in I} \|d^m \mathbf{u}(t)/dt^m\|$.]

Let $\boldsymbol{\theta}$ denote the zero vector in H. We introduce the following basic assumptions:

(A1) $\mathbf{f}(t, \boldsymbol{\theta}) = \boldsymbol{\theta}$ for all $t \in I$;

(A2) at any fixed $t \in I$, $\mathbf{f}(t, \cdot)$ has a Gateaux differential at every $\mathbf{u} \in U$ defined by

$$\mathbf{Df}(t, \mathbf{u}; \mathbf{h}) = \lim_{\lambda \to 0} [\mathbf{f}(t, \mathbf{u} + \lambda \mathbf{h}) - \mathbf{f}(t, \mathbf{u})]/\lambda,$$

where the limit exists for all $\mathbf{h} \in U$ in the sense of the norm of H. Condition (A1) implies that the trivial solution $\mathbf{u}(t) = \boldsymbol{\theta}$ on I is an equilibrium solution of (1). Note that condition (A2) may be satisfied when $\mathbf{f}$ involves differential operators. Thus, partial differential equations are not precluded here.

Let $\mathbf{u}(t; \mathbf{u}_0, t_0)$ denote a solution at t with initial data $\mathbf{u}_0$ at t_0.

Definition 1: The trivial solution of (1) is said to be

(i) *stable* if, for any $\epsilon > 0$, there exists a $\delta > 0$ such that $\|\mathbf{u}_0\| < \delta$ implies $\|\mathbf{u}(t; \mathbf{u}_0, t_0)\| < \epsilon$ for all $t \geq t_0$;

(ii) *asymptotically stable* (*locally*), if it is stable and there exists a $\bar{\delta} > 0$ such that $\|\mathbf{u}_0\| < \bar{\delta}$ implies $\|\mathbf{u}(t; \mathbf{u}_0, t_0)\| \to 0$ as $t \to \infty$;

(iii) *globally asymptotically stable* (relative to U), if it is asymptotically stable for any $\mathbf{u}_0 \in U$;

(iv) *explosively unstable* (relative to U), if for every nonzero $\mathbf{u}_0 \in U$, $\|\mathbf{u}(t; \mathbf{u}_0, t_0)\| \to \infty$ as $t \to t_1^- > t_0$ for some $t_1 \leq \infty$, where t_1 is the *explosion time* or *escape time*.

III. STABILITY AND INSTABILITY CONDITIONS

First, we give certain identities for abstract functions which will be used in establishing the main results.

Lemma 1: Let $\mathbf{g}$ be a mapping on a convex domain $D_g \subseteq H$ into H having a Gateaux differential at each point in D_g, then for every $\hat{z}, \hat{z} + z \in D_g$, we have

$$(\mathbf{g}(\hat{z} + z) - \mathbf{g}(\hat{z}), \mathbf{h}) = (\mathbf{Dg}(\hat{z} + \lambda z; z), \mathbf{h}) \tag{3}$$

for any $\mathbf{h} \in H$ and some $\lambda \in]0, 1[$, where λ depends on $\mathbf{h}$.

The above result is the Lagrange formula for operators on Banach spaces specialized to the case of a Hilbert space. A proof is given in Ref. 9. In the case where $\mathbf{g}$ is a Fréchet differentiable mapping, we have the following identity relating $\mathbf{g}$ and its Fréchet derivative $\mathbf{g}'$.

Lemma 2: Let $\mathbf{g}$ be a mapping on a convex domain

$D \subseteq H$ into H having a Fréchet derivative $g'(z)$ at every $z \in D_g$. Then

$$g(z) - g(\hat{z}) = \int_0^1 g'(\hat{z} + \lambda(z - \hat{z}))(z - \hat{z}) \, d\lambda \qquad (4)$$

for $z, \hat{z} \in D_g$.

Proof: Consider $g(p(\lambda))$, where p is a differentiable function of a real argument λ with $p(\lambda) \in D_g$. Since g is Fréchet differentiable on D_g, we have

$$(d/d\lambda) g(p(\lambda)) = g'(p(\lambda)) \, dp/d\lambda \qquad (5)$$

Let $p(\lambda) = \lambda z + (1 - \lambda)\hat{z}$ with $z, \hat{z} \in D_g$ and $\lambda \in [0, 1]$. Substituting this $p(\lambda)$ into (5) and integrating both sides lead directly to the desired identity (4). | |

Before considering the stability problem, we mention that the existence of a unique solution for the initial-value problem (1), (2) can be established by imposing suitable additional assumptions on f.[1,2] In particular, if the Fréchet derivative of f exists and is uniformly bounded on $I \times U$ (i.e., there exists a positive constant $M < \infty$ such that $\|f'(t, u)\| \leq M$ for all $(t, u) \in I \times U$), then, in view of Lemma 2, f satisfies a Lipschitz condition $\|f(t, u) - f(t, \hat{u})\|$ for any fixed $t \in I$ and all $u, \hat{u} \in U$. Moreover, if f is continuous in t on I for each fixed $u \in U$, then f is continuous on $I \times U$. Consequently, we may apply the classical Picard's method of successive approximations to establish the existence of a unique solution to (1) and (2).[1]

Now, we give the definitions for various types of mappings and results relating the properties of a mapping and its Gateaux differential.

Definition 2: A mapping g from its domain $D_g \subseteq H$ into H is said to be

(i) *coercive*, if $\mathrm{Re}(g(z), z) \geq \mu(\|z\|)\|z\|$ for all $z \in D_g$, where $\mu = \mu(s)$ is a real-valued function defined on $[0, \infty[$ such that $\mu(s) \to \infty$ as $s \to \infty$;

(ii) *monotone*, if $\mathrm{Re}(g(z) - g(\hat{z}), z - \hat{z}) \geq 0$ for all z, $\hat{z} \in D_g$;

(iii) *strongly monotone*, if $\mathrm{Re}(g(z) - g(\hat{z}), z - \hat{z}) \geq \gamma(\|z - \hat{z}\|)\|z - \hat{z}\|$ for any z, $\hat{z} \in D_g$, where $\gamma = \gamma(s)$ is a real-valued nonnegative function defined for $s \geq 0$ such that $\gamma(s) \to \infty$ as $s \to \infty$ and $\gamma(s) = 0$ if and only if $s = 0$.

(iv) *dissipative* (resp. *strongly dissipative*), if $-g$ is monotone (resp. strongly monotone).

Note that when g is a linear map, it is monotone (resp. strongly monotone) if $\mathrm{Re}(g(z), z) \geq 0$ (resp. $\mathrm{Re}(g(z), z) \geq \gamma(\|z\|)\|z\|$) for any $z \in D_g$. Also, when dealing with a family of mappings $\{g_\alpha : \alpha \in A\}$ having a common domain $D \subseteq H$ (for any fixed parameter α in a given set A, g_α is a mapping on D into H), we may define *uniform* coercivity, monotonicity, or dissipativity of $\{g_\alpha\}$ with respect to (abbrev. w.r.t.) A in an obvious way. Here, we require the functions μ and γ in Definitions 2(i) and 2(iii) respectively to be independent of α.

Lemma 3: For any fixed $t \in I$, let $g(t, \cdot)$ be a mapping with domain D (a linear subspace of H) and range in H, having a Gateaux differential $Dg(t, z; h)$ at any $z \in D$,

and $g(t, \theta) = \theta$. If $\{Dg(t, \tilde{z}; \cdot) : \tilde{z} \in D\}$ is uniformly coercive w.r.t. D for every $t \in I$, then $g(t, \cdot)$ is coercive for each $t \subseteq I$.

Proof: From Lemma 1 and $g(t, \theta) = \theta$ for each $t \in I$, we have, for every $t \in I$,

$$\mathrm{Re}(g(t, z), h) = \mathrm{Re}(Dg(t, \lambda z; h), h) \qquad (6)$$

at any $z, h \in D$ and some $\lambda \in \,]0, 1[$, where λ depends on h. Since $\{Dg(t, \tilde{z}; \cdot) : \tilde{z} \in D\}$ is uniformly coercive w.r.t. D, there exists a real-valued function $\mu = \mu(t, s)$, (independent of $\tilde{z}$) defined on $I \times [0, \infty[$ such that $\mu(t, s) \to \infty$ as $s \to \infty$ for any $t \in I$, and

$$\mathrm{Re}(Dg(t, \tilde{z}; h), h) \geq \mu(t, \|h\|)\|h\| \qquad (7)$$

for any fixed $(t, \tilde{z}) \in I \times D$ and all $h \in D$. Setting $\tilde{z} = \lambda z$ and $h = z$ in (6) and (7) leads to $\mathrm{Re}(g(t, z), z) \geq \mu(t, \|z\|)\|z\|$ for all $z \in D$, or $g(t, \cdot)$ is coercive. | |

Lemma 4: For any fixed $t \in I$, let $g(t, \cdot)$ be a mapping as in Lemma 3. If $\mathrm{Re}(Dg(t, z; h), h) \leq 0$ for each $t \in D$, then $\mathrm{Re}(g(t, z), h) \leq 0$ for all z, $h \in D$.

Proof: The desired result is evident from (6). | |

Remarks: (R1) In Lemmas 3 and 4, if $Dg(t, z; \cdot)$ is a linear (not necessarily continuous) map on D for each $(t, z) \in I \times D$, then we have the following specialized results:

(a) If $\{Dg(t, z; \cdot) : z \in D\}$ is uniformly strongly monotone [resp. uniformly strongly dissipative] w.r.t. D, then $g(t, \cdot)$ [resp. $-g(t, \cdot)$] is strongly monotone for each $t \in I$.

(b) If, for any fixed $t \in I$, $\{Dg(t, \tilde{z}; \cdot) : \tilde{z} \in D\}$ is uniformly dissipative w.r.t. D, then $\mathrm{Re}(g(t, z), h) \leq 0$ for each $(t, z) \in I \times D$ and all $h \in D$.

(R2) If, for any fixed $t \in I$, $g(t, \cdot)$ has a Fréchet derivative $g'(t, z)$ at any $z \in D$, we have from Lemma 2 and $g(t, \theta) = \theta$ the identity

$$g(t, z) = \int_0^1 g'(t, \lambda z) z \, d\lambda, \qquad (8)$$

where for any fixed $(t, \lambda z) \in I \times D$, $g'(t, \lambda z)$ is a bounded linear map on D into D. Moreover, if there exists a real-valued positive function $\beta = \beta(t)$ defined on I such that

$$\mathrm{Re}(g'(t, \tilde{z}) h, h) \geq \beta(t)\|h\|^2 \qquad (9)$$

for any $(t, \tilde{z}) \in I \times D$ and all $h \in D$, then, from (8), we have

$$\mathrm{Re}(g(t, z), z) \geq \beta(t)\|z\|^2 \qquad (10)$$

for all $(t, z) \in I \times D$.

Theorem 1: Let assumptions (A1), (A2) be satisfied. If (i) $\mathrm{Re}(Df(t, u; v), v) \leq 0$ for each $(t, u) \in I \times U$ and all $v \in U$; or (ii) $Df(t, u; \cdot)$ is a linear (not necessarily bounded) map on U into U for each $(t, u) \in I \times U$, and $\{Df(t, u; \cdot) : u \in U\}$ is uniformly dissipative w.r.t. D for each $t \in I$, then the trivial solution of (1) is stable.

Proof: Let $u(t) = u(t; u_0, t_0)$ be any solution of (1). Under assumptions (A1), (A2), and condition (i), it follows from Lemma 4 that

$$\frac{d}{dt}\|\mathbf{u}(t)\|^2 = 2\,\mathrm{Re}(\mathbf{f}(t,\mathbf{u}(t)),\mathbf{u}(t)) \leqslant 0 \tag{11}$$

with $\|\mathbf{u}(t_0)\| = \|\mathbf{u}_0\|$, which implies $\|\mathbf{u}(t)\| \leqslant \|\mathbf{u}_0\|$ for all $t \geqslant t_0$ or the stability of the trivial solution of (1). From Remark (R1)(b), estimate (11) remains valid under condition (ii). | |

Theorem 2: Assume that conditions (A1) and (A2) are satisfied. Moreover, there exists a continuous real-valued function $\mu = \mu(t,s)$ defined on $I \times [0,\infty[$ such that, for any $t \in I$, $\mu(t,s) = 0$ if and only if $s = 0$, and $\mu(t,s) \to \infty$ as $s \to \infty$, and

$$\mathrm{Re}(\mathbf{Df}(t,\mathbf{u};\mathbf{v}),\mathbf{v}) \leqslant -\mu(t,\|\mathbf{v}\|)\|\mathbf{v}\| \tag{12}$$

for any $(t,\mathbf{u}) \in I \times U$ and all $\mathbf{v} \in U$. Then, the asymptotic (local or global) stability of the trivial solution of the scalar differential equation

$$\frac{d\rho}{dt} = -\mu(t,\rho), \quad \rho(t_0) = \rho_0 \tag{13}$$

implies the asymptotic (local or global) stability of the trivial solution of (1).

Proof: Condition (12) implies that $\{-\mathbf{Df}(t,\mathbf{u};\circ):\mathbf{u} \in U\}$ is uniformly coercive w.r.t. U for each $t \in I$. We observe that under the hypotheses of the theorem, the trivial solution of (1) is the only equilibrium solution in U. To verify this, suppose there exists a nonzero $\tilde{\mathbf{u}} \in U$ such that $\mathbf{f}(t,\tilde{\mathbf{u}}) = \boldsymbol{\theta}$ for all $t \in I$, then, in view of (A1), (A2), Lemma 3, and Remark (R1)(a), we have

$$0 = \mathrm{Re}(\mathbf{f}(t,\tilde{\mathbf{u}}),\tilde{\mathbf{u}}) = \mathrm{Re}(\mathbf{Df}(\lambda\tilde{\mathbf{u}};\tilde{\mathbf{u}}),\tilde{\mathbf{u}}) \leqslant -\mu(t,\|\tilde{\mathbf{u}}\|)\|\tilde{\mathbf{u}}\| \tag{14}$$

for each $t \in I$ and some $\lambda \in]0,1[$. Since the upper bound in (14) is strictly negative, we have a contradiction which establishes the assertion.

Now, for any solution $\mathbf{u}(t) = \mathbf{u}(t;\mathbf{u}_0,t_0)$ of (1), we have, from (A1), (A2), (12), Lemma 3, and Remark (R1)(a), the estimate

$$\frac{d}{dt}\|\mathbf{u}(t)\|^2 = 2\,\mathrm{Re}(\mathbf{f}(t,\mathbf{u}(t)),\mathbf{u}(t)) \leqslant -2\mu(t,\|\mathbf{u}(t)\|)\|\mathbf{u}(t)\|,$$

or

$$\frac{d}{dt}\|\mathbf{u}(t)\| \leqslant -\mu(t,\|\mathbf{u}(t)\|) \leqslant 0 \tag{15}$$

with $\|\mathbf{u}(t_0)\| = \|\mathbf{u}_0\|$. Thus, we have $\|\mathbf{u}(t)\| \leqslant \|\mathbf{u}_0\|$ for all $t \geqslant t_0$ or the stability of the trivial solution of (1).

Let $\hat{\rho}(t;\rho_0,t_0)$ be the maximal solution of (13) defined on some interval $[t_0,t_1] \subseteq I$. Such a solution exists when μ is continuous on $I \times [0,\infty[$.[10] By setting $\|\mathbf{u}_0\| = \rho_0$, we have the estimate

$$\|\mathbf{u}(t;\mathbf{u}_0,t_0)\| \leqslant \hat{\rho}(t;\|\mathbf{u}_0\|,t_0) \quad \text{for all } t \in [t_0,t_1]. \tag{16}$$

Since the trivial solution of (13) is asymptotically (locally or globally) stable, estimate (16) is valid for all $t \in I$. Now, the desired result is evident from (16). For the case of global asymptotic stability, the solution initiated from any $\mathbf{u}_0 \in U$ at t_0 tends to the unique equilibrium point $\boldsymbol{\theta}$ as $t \to \infty$. | |

Theorem 3: Let assumptions (A1) and (A2) be satisfied. Also, $\mathbf{Df}(t,\mathbf{u};\cdot)$ is a linear (not necessarily bounded) transformation on U into U for each $(t,\mathbf{u}) \in I$

$\times U$, and there exists a continuous real-valued positive function $\gamma = \gamma(t)$ defined on I such that

$$\mathrm{Re}(\mathbf{Df}(t,\mathbf{u};\mathbf{v}),\mathbf{v}) \leqslant -\gamma(t)\|\mathbf{v}\|^2 \tag{17}$$

for each $(t,\mathbf{u}) \in I \times U$ and all $\mathbf{v} \in U$. Then the solutions of (1), (2) satisfy the estimate

$$\|\mathbf{u}(t;\mathbf{u}_0,t_0)\| \leqslant \|\mathbf{u}_0\| \exp\left(\int_{t_0}^{t} -\gamma(s)\,ds\right) \quad \text{for any } t \in I. \tag{18}$$

Furthermore, if

$$\limsup_{t \to \infty} \int_{t_0}^{t} \gamma(s)\,ds = +\infty, \tag{19}$$

then the trivial solution of (1) is globally asymptotically stable.

Proof: Condition (17) implies that $\{\mathbf{Df}(t,\mathbf{u};\circ):\mathbf{u} \in U\}$ is uniformly strongly dissipative w.r.t. U for each $t \in I$. From Remark (R1)(a), for each $t \in I$, $\mathbf{f}(t,\cdot)$ is strongly dissipative and satisfies the estimate

$$\mathrm{Re}(\mathbf{f}(t,\mathbf{u}),\mathbf{u}) \leqslant -\gamma(t)\|\mathbf{u}\|^2 \tag{20}$$

for all $\mathbf{u} \in U$. Thus, along any solution $\mathbf{u}(t) = \mathbf{u}(t;\mathbf{u}_0,t_0)$,

$$\frac{d\|\mathbf{u}(t)\|}{dt} \leqslant -\gamma(t)\|\mathbf{u}(t)\| \quad \text{for all } t \in I, \tag{21}$$

with $\|\mathbf{u}(t_0)\| = \|\mathbf{u}_0\|$, which leads to (18). As in Theorem 2, the trivial solution is the only equilibrium solution of (1). Its global asymptotic stability under condition (19) is evident from (18). | |

Theorem 4: Let assumptions (A1) and (A2) be satisfied. If there exists a real-valued function $\mu = \mu(t,s)$ which is continuous on $I \times [0,\infty[$ and $\mu(t,s) \to \infty$ as $s \to \infty$ for each $t \in I$ such that

$$\mathrm{Re}(\mathbf{Df}(t,\mathbf{u};\mathbf{v}),\mathbf{v}) \geqslant \mu(t,\|\mathbf{v}\|)\|\mathbf{v}\| \tag{22}$$

for every $(t,\mathbf{u}) \in I \times U$ and all $\mathbf{v} \in U$. Then, the explosive instability of the trivial solution of the scalar differential equation:

$$\frac{d\rho}{dt} = \mu(t,\rho), \quad \rho(t_0) = \rho_0 \tag{23}$$

implies the explosive instability of the trivial solution of (1).

Proof: Under the conditions of the theorem, we have from Lemma 3 the estimate

$$\frac{d\|\mathbf{u}(t)\|}{dt} \geqslant \mu(t,\|\mathbf{u}(t)\|) \tag{24}$$

along any solution $\mathbf{u}(t) = \mathbf{u}(t;\mathbf{u}_0,t_0)$ with $\|\mathbf{u}(t_0)\| = \|\mathbf{u}_0\|$. Since μ is continuous on $I \times [0,\infty[$, there exists a minimal solution $\check{\rho}(t;\rho_0,t_0)$ of (20) defined on some interval $[t_0,t_1] \subseteq I$ [10, p. 16]. Setting $\|\mathbf{u}_0\| = \rho_0$ leads to the estimate

$$\|\mathbf{u}(t;\mathbf{u}_0,t_0)\| \geqslant \check{\rho}(t;\|\mathbf{u}_0\|,t_0) \quad \text{for all } t \in [t_0,t_1], \tag{25}$$

from which the desired result follows. | |

Remarks: (R3) A result similar to Theorem 4 can be readily established for the case where condition (22) is replaced by $\mathrm{Re}(\mathbf{Df}(t,\mathbf{u};\mathbf{v}),\mathbf{v}) \geqslant \gamma(t)\|\mathbf{v}\|^2$; in this case we have the lower bound

$$\|\mathbf{u}(t;\mathbf{u}_0,t_0)\| \geqslant \|\mathbf{u}_0\| \exp\left(\int_{t_0}^{t} \gamma(s)\,ds\right), \quad t \geqslant t_0. \tag{26}$$

Evidently, under condition (19), we have explosive instability with explosion time $t_1 \leqslant \infty$.

(R4) If $f(t, \circ)$ is a linear mapping on U into H for each $t \in I$, then its Gateaux derivative coincides with $f(t, \circ)$, and $(Df(t, u; v), v) = (f(t, u), v)$ for all $u \in D_t$ and $v \in H$. Thus, the foregoing stability and instability conditions reduce to the usual ones.

IV. APPLICATIONS

In this section, we apply the main results developed in Sec. III to specific forms of nonlinear evolutional equations which arise in physics.

1 Consider a system of ordinary differential equations of the form (1) defined on the complex n-dimensional space $\mathbb{C}^n$ with inner product defined by

$$(u, v) = \sum_{i=1}^{n} u_i \bar{v}_i \tag{27}$$

for any $u, v \in \mathbb{C}^n$, where $(\bar{\cdot})$ denotes complex conjugation. Let f be a continuous function on $I \times \mathbb{C}^n$ into $\mathbb{C}^n$ satisfying assumptions (A1) and (A2). If we assume further that, for each fixed $t \in I$, $f(t, u)$ has continuous first partial derivatives with respect to u_i on $\mathbb{C}^n$, then $f(t, \cdot)$ has a Fréchet derivative everywhere on $\mathbb{C}^n$ given by $f'(t, u) = J_t(t, u)$, where $J_t(t, u)$ is the Jacobian matrix of $f(t, \cdot)$ at u. Thus, we can write

$$\mathrm{Re}(f'(t, u)v, v) = \tfrac{1}{2}(Q(t, u)v, v), \tag{28}$$

where $Q(t, u)$ is the Hermitian matrix given by

$$Q(t, u) = J_t^*(t, u) + J_t(t, u), \tag{29}$$

where $(\circ)^*$ denotes conjugate transposition. Let $\lambda_{max}(t, u)$ and $\lambda_{min}(t, u)$ denote the maximum and minimum eigenvalues of $Q(t, u)$ respectively. Assume that the following quantities exist:

$$\check{\lambda}(t) = \inf_{u \in \mathbb{C}^n} \lambda_{min}(t, u), \quad \hat{\lambda}(t) = \sup_{u \in \mathbb{C}^n} \lambda_{max}(t, u). \tag{30}$$

From Theorem 3, if

$$\limsup_{t \to \infty} \int_{t_0}^{t} \hat{\lambda}(s)\, ds = -\infty, \tag{31}$$

then the trivial solution is globally asymptotically stable. Note that if we take H to be the real n-dimensional Euclidean space $\mathbb{R}^n$ with the usual inner product, and assume that $Q(t, u)$ is uniformly negative definite on $I \times \mathbb{R}^n$, then the foregoing result reduces to that of Krasovskii.[11]

Also, from Theorem 4 and Remark (R3), if

$$\liminf_{t \to \infty} \int_{t_0}^{t} \check{\lambda}(s)\, ds = +\infty, \tag{32}$$

then the trivial solution is explosively unstable.

The foregoing results can be generalized by defining a new inner product on $\mathbb{C}^n$:

$$(u, v)_P = (u, Pv), \tag{33}$$

where P is any positive-definite $n \times n$ Hermitian matrix. In this case, we can write

$$\mathrm{Re}(f'(t, u)v, v)_P = \tfrac{1}{2}(\bar{Q}(t, u)v, v), \tag{34}$$

where $\tilde{Q}(t, u)$ is the Hermitian matrix given by

$$\bar{Q}(t, u) = J_t^*(t, u)P + PJ_t(t, u). \tag{35}$$

Replacing $Q(t, u)$ by $\bar{Q}(t, u)$, we obtain stability and instability conditions corresponding to (31) and (32) respectively. In the special case where $H = \mathbb{R}^n$ and $\tilde{Q}(t, u)$ is uniformly negative definite on $I \times \mathbb{R}^n$, we have the generalized Krasovskii's theorem for a nonautonomous system.[12] Also, when f is a time-independent continuously differentiable function on $\mathbb{R}^n$ into $\mathbb{R}^n$ such that for some positive-definite real symmetric matrix P, the maximum eigenvalue $\hat{\lambda}_{\tilde{Q}}(u)$ of $\bar{Q}(u) = J_t^T(u)P + PJ_t(u)$ satisfy

$$\hat{\lambda}_{\tilde{Q}}(u) < -\nu(\|u\|_P) \quad \text{for every } u \in \mathbb{R}^n, \tag{36}$$

where $\|u\|_P^2 = (u, u)_P$ and $\nu = \nu(s)$ is a real-valued continuous, monotone decreasing, positive function of s on $[0, \infty[$, then we have

$$\frac{d}{dt}\|u(t)\|_P^2 = 2(f(u(t)), Pu(t))$$

$$= 2 \int_0^1 (f'(\xi u(t))u(t), Pu(t))\, d\xi$$

$$= \int_0^1 (\bar{Q}(\xi u(t))u(t), u(t))\, d\xi \leqslant \|u(t)\|^2 \int_0^1 \hat{\lambda}_{\tilde{Q}}(\xi u(t))\, d\xi$$

$$\leqslant -\|u(t)\|^2 \int_0^1 \nu(\|\xi u(t)\|_P)\, d\xi$$

$$\leqslant -\hat{\lambda}_P^{-1} \|u(t)\|_P^2 \int_0^1 \nu(\xi \|u(t)\|_P)\, d\xi$$

$$\leqslant -\hat{\lambda}_P^{-1} \|u(t)\|_P^2 \nu(\|u(t)\|_P), \quad t > 0, \tag{37}$$

where $\hat{\lambda}_P$ is the maximum eigenvalue of P. Thus,

$$\frac{d}{dt}\|u(t)\|_P \leqslant -\hat{\lambda}_P^{-1} \|u(t)\|_P \nu(\|u(t)\|_P)/2 \leqslant 0, \quad t > 0. \tag{38}$$

Evidently, the asymptotic (local or global) stability of the trivial solution of the scalar differential equation

$$\frac{d\rho}{dt} = -\tfrac{1}{2}\hat{\lambda}_P^{-1}\rho\nu(\rho), \quad \rho(0) = \rho_0 \tag{39}$$

implies the asymptotic (local or global) stability of the trivial solution of $du/dt = f(u)$. This result is essentially that of Markus and Yamabe.[13]

Finally, we note that the evolutional equations defined on $\mathbb{C}^n$ arising from physical situations may not have Fréchet differentiable right-hand sides. For example, the equation for the complex amplitudes of a nonlinear three-wave interacting system has the form[14]

$$\frac{du}{dt} = A(t)u + g(t, u) \overset{\Delta}{=} f(t, u) \tag{40}$$

with

$$A(t) = \begin{bmatrix} i\omega_1(t) & & 0 \\ & i\omega_2(t) & \\ 0 & & i\omega_3(t) \end{bmatrix},$$

$$\mathbf{g}(t,\mathbf{u}) = \begin{bmatrix} \bar{\mu}_1(t)u_2u_3 \\ \mu_2(t)u_1\bar{u}_3 \\ \mu_3(t)u_1\bar{u}_2 \end{bmatrix} , \qquad (41)$$

where $i = \sqrt{-1}$; $\omega_j(t)$ corresponds to the complex frequency of the jth wave at time t, and μ_j's are the time-dependent complex coupling coefficients. Here, $\mathbf{f}(t, \cdot)$ has a Gateaux differential everywhere in $\mathbb{C}^3$ given by

$$\mathbf{Df}(t,\mathbf{u};\mathbf{v}) = \mathbf{A}(t)\mathbf{v} + \mathbf{Dg}(t,\mathbf{u};\mathbf{v}),$$

$$\mathbf{Dg}(t,\mathbf{u};\mathbf{v}) = \begin{bmatrix} \bar{\mu}_1(t)(u_2v_3 + v_2u_3) \\ \mu_2(t)(u_1\bar{v}_3 + v_1\bar{u}_3) \\ \mu_3(t)(u_1\bar{v}_2 + v_1\bar{u}_2) \end{bmatrix} . \qquad (42)$$

Evidently, for any fixed $(t,\mathbf{u}) \in I \times \mathbb{C}^3$, $\mathbf{Dg}(t,\mathbf{u}; \cdot)$ is *not* a linear transformation on $\mathbb{C}^3$. Hence $\mathbf{f}(t, \cdot)$ is not Fréchet differentiable on $\mathbb{C}^3$. From (42), we have

$$\mathrm{Re}(\mathbf{Df}(\lambda\mathbf{u};\mathbf{u}),\mathbf{u})$$

$$= -\sum_{j=1}^{3} \mathrm{Im}\,\omega_j(t)|u_j|^2 + 2\lambda\,\mathrm{Re}\{\sum_{j=1}^{3} \mu_j(t)u_1\bar{u}_2\bar{u}_3\}. \qquad (43)$$

By direct computation, we verify that (43) is equal to $\mathrm{Re}(\mathbf{f}(t,\mathbf{u}(t)),\mathbf{u}(t))$ when $\lambda = \frac{1}{2}$. It can be shown that if $|\mathrm{Im}\,\omega_j(t)|$ and $|\mu_j(t)|$ are bounded on I for $j = 1,2,3$, then there exist constants $\alpha^{\pm}$ and β such that

$$-(\alpha^{-}\|\mathbf{u}\|^2 + \beta\|\mathbf{u}\|^4) \le 2\,\mathrm{Re}(\mathbf{f}(t,\mathbf{u}),\mathbf{u}) \le \alpha^{+}\|\mathbf{u}\|^2 + \beta\|\mathbf{u}\|^4 \quad (44)$$

for all $t \in I$. These bounds may be used to obtain stability and instability conditions (see Ref. 15 for details). Of course, (40) can be reformulated as an evolutional equation on $\mathbb{R}^6$, whose right-hand side is Fréchet differentiable everywhere in $\mathbb{R}^6$.

2. Let $\tau \ge 0$ be a given real number and H be a complex Hilbert space with inner product $(\cdot , \cdot)$. Let $\mathbf{z} \in C_0([-\tau,\infty[;H)$. Fro any $t \in I = [0,\infty[$, we define $\mathbf{z}_t$ by

$$\mathbf{z}_t(s) = \mathbf{z}(t+s), \quad -\tau \le s \le 0. \qquad (45)$$

Now, consider the functional differential equation:

$$\frac{d\mathbf{z}(t)}{dt} = \mathbf{F}(\mathbf{z}_t), \qquad (46)$$

where $\mathbf{F}:C_0([-\tau,0];H) \to H$ is a given function satisfying $\mathbf{F}(\tilde{\theta}) = \theta_H$, where $\tilde{\theta}$ and θ_H are the zero vectors in $C_0([-\tau,0];H)$ and H respectively. The domain of $\mathbf{F}$ is taken to be a subset $D_{\mathbf{F}}$ of $C_0([-\tau,0];H)$ consisting of H-valued functions which are absolutely continuous w.r.t. t on $[-\tau,0]$. The initial data at $t=0$ for (46) is given by

$$\mathbf{z}_0(s) = \phi_0(s), \quad -\tau \le s \le 0, \qquad (47)$$

where $\phi_0 \in D_{\mathbf{F}}$. A function $\mathbf{z}:I \to H$ is said to be a solution to (46), (47), if $\mathbf{z}_t \in D_{\mathbf{F}}$ for each $t \in I$ with $\mathbf{z}_0 = \phi_0$. Moreover, $\mathbf{z}(t)$ is absolutely continuous w.r.t. t on I. {In general, a solution which is continuously differentiable w.r.t. t may not exist. For example, the scalar equation $\dot{z}(t) = \dot{z}(t-1) + z(t)$ with continuous initial data $z(s) = \phi_0(s)$, $s \in [-1,0]$ does not have a C_1 solution for $t \ge 0$ when ϕ_0 is C_1 on $[-1,0]$ but $\dot{\phi}_0(0) \ne \dot{\phi}_0(-1) + \phi_0(0)$.

As in Ref. 16, we reformulate the foregoing equation

as a nonlinear evolutional equation in the form of (1) in a suitable Hilbert space H. In the sequel, we take H to be $H \times L_2([\tau,0];H)$ whose inner product is defined by

$$(\mathbf{u},\mathbf{v}) = (\eta,\tilde{\eta})_H + \int_{-\tau}^{0} (\phi(s), \tilde{\phi}(s))_H \, ds \qquad (48)$$

for any $\mathbf{u} = (\eta,\phi)$, $\mathbf{v} = (\tilde{\eta},\tilde{\phi}) \in H$. Let $\mathbf{f}$ be the mapping from H into H defined by

$$\mathbf{f}(\mathbf{u}) = \begin{bmatrix} \mathbf{F}(\phi) \\ \phi' \end{bmatrix} , \quad \phi' = \frac{d\phi}{ds} \qquad (49)$$

with domain

$$D_{\mathbf{f}} = \{\mathbf{u} = (\eta,\phi) \in H : \phi \in D_{\mathbf{F}}, \ \phi' \in L_2([-\tau,0];H), \ \eta = \phi(0)\}. \qquad (50)$$

Here, a solution to the initial-value problem for the foregoing evolutional equation with initial data $\mathbf{u}(0) = (\phi_0(0), \phi_0(\cdot)) \in D_{\mathbf{f}}$ is defined as a function $\mathbf{u}:I \to H$ satisfying the given initial data such that for each $t \in I$, $\mathbf{u}(t) \in D_{\mathbf{f}}$ and $\mathbf{u}(t)$ is absolutely continuous w.r.t. t on I. Evidently, if $\mathbf{u}(t) = (\mathbf{z}(t),\mathbf{z}_t)$ is a solution to the initial-value problem for the evolutional equation, then $\mathbf{z}(t)$ is a solution to (46) with initial data ϕ_0. Moreover, the stability (or asymptotic stability) of the trivial solution of the evolutional equation w.r.t. the H-norm implies the stability (or asymptotic stability) of the trivial solution of (46) w.r.t. the norms for H and H. [The trivial solution of (1) with $\mathbf{f}$ defined by (49) is stable, if for every $\epsilon > 0$, there exists a $\delta > 0$ such that $\|\phi_0\| < \delta$ implies $\|\mathbf{z}(t;\phi_0,0)\|_H < \epsilon$ for all $t \ge 0$, where $\| \cdot \|$ and $\| \cdot \|_H$ denote the norms for H and H respectively.] In fact, the converse of this statement is also true, since for $\|\phi_0\| < \delta$ such that $\|\mathbf{z}(t;\phi_0,0)\|_H < \epsilon$ for all $t \ge 0$, we have $\|\mathbf{u}(t;\mathbf{u}_0,0)\| < (1+\tau)^{1/2}\epsilon$ for all $t \ge 0$.

Now, if $\mathbf{F}$ has a Gateaux differential on $D_{\mathbf{F}}$, then $\mathbf{f}$ has a Gateaux differential on $D_{\mathbf{f}}$ given by

$$\mathbf{Df}(\mathbf{u};\mathbf{v}) = \begin{bmatrix} \mathbf{DF}(\phi;\tilde{\phi}) \\ \tilde{\phi}' \end{bmatrix} , \qquad (51)$$

where $\mathbf{u} = (\eta,\phi)$, $\mathbf{v} = (\tilde{\eta},\tilde{\phi}) \in D_{\mathbf{f}}$. Thus,

$$\mathrm{Re}(\mathbf{Df}(\mathbf{u};\mathbf{v}),\mathbf{v}) = \mathrm{Re}(\mathbf{DF}(\phi;\tilde{\phi}), \tilde{\phi}(0))_H + \int_{-\tau}^{0} \mathrm{Re}(\tilde{\phi}'(s), \tilde{\phi}(s))_H \, ds$$

$$= \mathrm{Re}(\mathbf{DF}(\phi;\tilde{\phi}), \phi(0))_H + \tfrac{1}{2}(\|\tilde{\phi}(0)\|_H^2 - \|\tilde{\phi}(-\tau)\|_H^2). (52)$$

Suppose that a constant γ can be found such that

$$\mathrm{Re}(\mathbf{Df}(\mathbf{u};\mathbf{v}),\mathbf{v}) \le -\gamma\{\|\tilde{\phi}(0)\|_H^2 + \int_{-\tau}^{0} \|\tilde{\phi}(s)\|_H^2 \, ds\} \qquad (53)$$

for all $\mathbf{u},\mathbf{v} \in D_{\mathbf{f}}$. From Theorem 3 [resp. Theorem 1], if γ is positive [resp. zero], the foregoing evolutional equation is globally asymptotically stable (relative to $D_{\mathbf{f}}$) [resp. stable], which implies that the trivial solution of (46) is also globally asymptotically stable (relative to $D_{\mathbf{F}}$) [resp. stable].

Now, we apply the foregoing results to the following linear time-lag system defined on $\mathbb{C}^n$:

$$\frac{d\mathbf{z}(t)}{dt} = \mathbf{A}\mathbf{z}(t) + \mathbf{B}\mathbf{z}(t-\tau), \quad 0 \le \tau < \infty, \qquad (54)$$

where $\mathbf{A}$ and $\mathbf{B}$ are complex $n \times n$ matrices. The initial

data at $t=0$ is given as in (47) with $\phi_0 \in C_0([-\tau, 0]; \mathbb{C}^n)$. To reformulate (54) as an evolutional equation in $H = \mathbb{C}^n \times L_2([-\tau, 0]; \mathbb{C}^n)$ in the form of (1), we define $\mathbf{u} = (\eta, \phi)$ and

$$\mathbf{f}(\mathbf{u}) = \begin{bmatrix} \mathbf{A}\phi(0) + \mathbf{B}\phi(-\tau) \\ \phi' \end{bmatrix}, \quad \phi' = \frac{d\phi}{ds} \tag{55}$$

with domain D_f as defined by (50).

Now, by direct computation, we have for $\mathbf{v} = (\phi(0), \phi(\cdot)) \in D_f$:

$$\mathrm{Re}(\mathbf{Df}(\mathbf{u}; \mathbf{v}), \mathbf{v})$$

$$= \mathrm{Re}(\mathbf{A}\phi(0) + \mathbf{B}\phi(-\tau), \phi(0))_{\mathbb{C}^n} + \tfrac{1}{2}(\|\phi(0)\|^2_{\mathbb{C}^n} - \|\phi(-\tau)\|^2_{\mathbb{C}^n})$$

$$= \tfrac{1}{2}\{(\phi(0), (\mathbf{I} + \mathbf{A} + \mathbf{A}^*)\phi(0))_{\mathbb{C}^n} + (\phi(0), (\mathbf{B} + \mathbf{B}^*)\phi(-\tau))_{\mathbb{C}^n}$$

$$- (\phi(-\tau), \phi(-\tau))_{\mathbb{C}^n}\} = \tfrac{1}{2}(\mathbf{Q}\mathbf{w}, \mathbf{w})_{\mathbb{C}^{2n}}, \tag{56}$$

where $\mathbf{w} = (\phi(0), \phi(-\tau))$ and

$$\mathbf{Q} = \begin{bmatrix} \mathbf{I} + \mathbf{A} + \mathbf{A}^* & (\mathbf{B} + \mathbf{B}^*)/2 \\ (\mathbf{B} + \mathbf{B}^*)/2 & -\mathbf{I} \end{bmatrix}. \tag{57}$$

Evidently, if $\mathbf{Q}$ is negative definite, then $\mathrm{Re}(\mathbf{Df}(\mathbf{u}; \mathbf{v}), \mathbf{v}) \leq 0$ for all $\mathbf{u}, \mathbf{v} \in D_f$, and we have stability of the trivial solution of (54). For the case of a single complex equation (54) with $A = a$ and $B = b$, the matrix $\mathbf{Q}$ is negative definite, if the Sylvester inequalities

$$2\,\mathrm{Re}(a) + 1 < 0 \quad \text{and} \quad [1 + 2\mathrm{Re}(a)] + \mathrm{Re}(b)]^2 < 0$$

are satisfied. For real a and b, the above conditions may be compared with the following ones for asymptotic stability as deduced from Pontryagin's results pertaining to the zeros of transcendental functions [17,18]:

$$\tau a < 1 \quad \text{and} \quad \tau a > b\tau > -(\gamma_0^2 + \tau^2 a^2)^{1/2}, \tag{59}$$

where γ_0 is the root $\gamma = a\tan\gamma$ for $0 < \gamma < \pi$; and for $a = 0$, we take $\gamma_0 = \pi/2$. It can be readily verified that the region in the (a, b) plane defined by (58) is a proper subset of that defined by (59). Note that (58) is independent of the delay time τ. Hence, the negative definiteness of $\mathbf{Q}$ ensures stability of the trivial solution for all $\tau \geq 0$.

3. Consider a system of complex partial differential equations of the form:

$$\frac{\partial \mathbf{u}(t, \mathbf{x})}{\partial t} = [\mathbf{A}(t, \cdot)\mathbf{u}(t, \cdot)](\mathbf{x}) + \mathbf{g}(t, \mathbf{x}, \mathbf{u}(t, \mathbf{x})) \tag{60}$$

defined for $t > t_0$ and $\mathbf{x} = (x_1, \ldots, x_m) \in \Omega$—an open connected subset of $\mathbb{R}^m$, where $\mathbf{u} = (u_1, \ldots, u_n)$, $\mathbf{A}(t, \cdot)$ is a linear operator for each $t \in I$, and $\mathbf{g}: I \times \Omega \times \mathbb{C}^n \to \mathbb{C}^n$ is a continuous mapping which is continuous differentiable w.r.t. $\mathbf{u}$ for any fixed $(t, \mathbf{x}) \in I \times \bar{\Omega}$ ($\bar{\Omega}$ denotes the closure of Ω). Moreover, $\mathbf{g}(t, \mathbf{x}, 0) = 0$ for all $(t, \mathbf{x}) \in I \times \bar{\Omega}$.

Let H be a suitable complex Hilbert space of $\mathbb{C}^n$-valued functions. If for any fixed $t \in I$, the domain of $\mathbf{A}(t) \equiv \mathbf{A}(t, \cdot)$ can be chosen to be a time-independent linear subspace $D_\mathbf{A}$ of H such that $\mathbf{A}(t, \cdot)$ can be chosen to be a time-independent linear subspace $D_\mathbf{A}$ of H such that $\mathbf{A}(t, \cdot)\mathbf{u} \in D_\mathbf{A}$; and $\mathbf{g}(t, \cdot, \mathbf{u}) \in H$ for any $\mathbf{u} \in H$, we may reformulate (60) as an evolutional equation in H of the form:

$$\frac{d\mathbf{u}}{dt} = \mathbf{A}(t)\mathbf{u} + \mathbf{g}(t, \mathbf{u}) \triangleq \mathbf{f}(t, \mathbf{u}). \tag{61}$$

For this system, $\mathbf{f}$ has a Gateaux differential given by

$$(\mathbf{Df}(t, \mathbf{u}; \mathbf{v}), \mathbf{v}) = (\mathbf{A}(t)\mathbf{v}, \mathbf{v}) + (\mathbf{J}_\mathbf{g}(t, \cdot; \mathbf{u})\mathbf{v}, \mathbf{v}), \tag{62}$$

where $\mathbf{J}_\mathbf{g}(t, \mathbf{x}; \mathbf{u})$ is the Jacobian matrix of $\mathbf{g}$ w.r.t. $\mathbf{u}$ at $(t, \mathbf{x}, \mathbf{u})$ and $\mathbf{v} \in D_\mathbf{A}$. Now, if there exist real-valued functions $\check{\gamma}_i$ and $\hat{\gamma}_i$, $i = 1, 2$, defined on I such that for each $t \in I$ and all $\mathbf{v} \in D_\mathbf{A}$

$$\check{\gamma}_1(t)\|\mathbf{v}\|^2 \leq \mathrm{Re}(\mathbf{A}(t)\mathbf{v}, \mathbf{v}) \leq \hat{\gamma}_1(t)\|\mathbf{v}\|^2 \tag{63}$$

and

$$\check{\gamma}_2(t)\|\mathbf{v}\|^2 \leq \mathrm{Re}(\mathbf{J}_\mathbf{g}(t, \cdot; \mathbf{u})\mathbf{v}, \mathbf{v}) \leq \hat{\gamma}_2(t)\|\mathbf{v}\|^2 \quad \text{for all } \mathbf{u} \in H, \tag{64}$$

then we may apply Theorem 3 or 4 to obtain stability or instability conditions for the trivial solution of this system. For specificness, we consider a few particular systems arising from physics.

a: Diffusion System: Let Ω be a bounded open connected subset of $\mathbb{R}^m$ with boundary $\partial\Omega$, and H be the complex Hilbert space $L_2^n(\Omega)$ [n-fold Cartesian product of $L_2(\Omega)$] with inner product

$$(\mathbf{u}, \mathbf{v}) = \int_\Omega \sum_{i=1}^n u_i(x)\bar{v}_i(x)\,d\Omega \quad \text{for } \mathbf{u}, \bar{\mathbf{v}} \in H. \tag{65}$$

The operator $\mathbf{A}(t)$ has the form

$$\mathbf{A}(t)\mathbf{u} = \mathbf{A}_0(t)\mathbf{u} + \sum_{j=1}^m \mathbf{A}_{1j}(t, \mathbf{x})\frac{\partial \mathbf{u}}{\partial x_j}, \tag{66}$$

where $\mathbf{A}_{1j}(t, \mathbf{x})$, $j = 1, \ldots, m$, are $n \times n$ Hermitian matrices whose elements are continuous functions on $I \times \bar{\Omega}$ and, for each $t \in I$, they are continuously differentiable w.r.t. $\mathbf{x}$ on $\bar{\Omega}$. $\mathbf{A}_0(t)$ is an $n \times n$ diagonal matrix operator whose diagonal elements $\mathbf{A}_{0i}(t)$ are second-order elliptic operators given by

$$\mathbf{A}_{0i}(t)u_i = \sum_{j,k=1}^m \frac{\partial}{\partial x_j} a_{jk}^{(i)}(t, \mathbf{x})\frac{\partial u_i}{\partial k_k}, \quad i = 1, \ldots, n, \tag{67}$$

where, for each fixed $i = 1, \ldots, n$, the coefficients $a_{jk}^{(i)}$ are continuous functions of t and $\mathbf{x}$ on $I \times \bar{\Omega}$. Moreover, they satisfy

$$\sum_{j,k=1}^m a_{jk}^{(i)}(t, \mathbf{x})\xi_j\bar{\xi}_k \geq \sigma_i(t)\sum_{j=1}^m |\xi_j|^2 \tag{68}$$

for every $\mathbf{x} \in \bar{\Omega}$ and $\xi = (\xi_1, \ldots, \xi_m) \in \mathbb{C}^m$, where σ_i is a real continuous positive function of t on I.

At the boundary $\partial\Omega$, we impose the condition

$$\mathbf{u}(t, \mathbf{x}) = 0 \quad \text{for all } (t, \mathbf{x}) \in I \times \partial\Omega. \tag{69}$$

For this system, we may take $D_\mathbf{A}$ to be the linear subspace of $L_2^n(\Omega)$ consisting of all continuous functions which vanish on $\partial\Omega$ and have continuous partial derivatives w.r.t. x_j up the second order. This type of systems arises in the study of the simultaneous diffusion of several substances such as the multigroup neutrons in a nuclear reactor and a multispecies collision-dominated plasma.

Assuming that $\partial\Omega$ is sufficiently smooth, we may integrate by parts and apply Green's theorem to obtain

$$\text{Re}(\mathbf{A}(t)\mathbf{v}, \mathbf{v}) = -\int_\Omega \sum_{i=1}^{n} \sum_{j,k=1}^{m} a_{jk}^{(i)} \frac{\partial v_i}{\partial x_j} \frac{\partial \bar{v}_i}{\partial x_k} \, d\Omega$$

$$-\frac{1}{2}\sum_{j=1}^{m} \left(\frac{\partial \mathbf{A}_{1j}}{\partial x_j} \mathbf{v}, \mathbf{v} \right) \quad \text{for all } \mathbf{v} \in \mathcal{D}_\mathbf{A}. \qquad (70)$$

Using (68) and Poincaré's inequality[19] for each v_i

$$\int_\Omega \sum_{j=1}^{m} \left| \frac{\partial v_i}{\partial x_j} \right|^2 d\Omega \geqslant \lambda \int_\Omega |v_i|^2 \, d\Omega, \qquad (71)$$

leads to

$$\text{Re}(\mathbf{A}(t)\mathbf{v}, \mathbf{v}) \leqslant -\left(\left[\lambda \min_i \{\sigma_i(t)\} \mathbf{I}_H + \frac{1}{2}\sum_{j=1}^{m} \frac{\partial \mathbf{A}_{1j}}{\partial x_j} \right] \mathbf{v}, \mathbf{v} \right), \qquad (72)$$

where $\mathbf{I}_H$ is the identity operator on H and λ is a positive constant depending only on Ω.

From (62) and (72), we have

$$\text{Re}(\mathbf{Df}(t,\mathbf{u};\mathbf{v}), \mathbf{v}) \leqslant -([\lambda \min_i \{\sigma_i(t)\} \mathbf{I}_H + \mathbf{Q}(t, \circ ;\mathbf{u})\mathbf{v}, \mathbf{v}) \qquad (73)$$

for each $(t, \mathbf{u}) \in I \times \mathcal{D}_\mathbf{A}$ and all $\mathbf{v} \in \mathcal{D}_\mathbf{A}$, where

$$\mathbf{Q}(t, \cdot ;\mathbf{u}) = \frac{1}{2} \left\{ \sum_{j=1}^{m} \frac{\partial \mathbf{A}_{1j}(t,\mathbf{x})}{\partial x_j} - \mathbf{J}_g^*(t,\mathbf{x};\mathbf{u}) - \mathbf{J}_g(t,\mathbf{x};\mathbf{u}) \right\}. \qquad (74)$$

Now, if there exists a real-valued continuous positive function $\alpha = \alpha(t)$ defined on I such that

$$(\mathbf{Q}(t, \cdot ;\mathbf{u})\mathbf{v}, \mathbf{v}) \geqslant \alpha(t) \|\mathbf{v}\|^2 \qquad (75)$$

for every fixed $(t, \mathbf{u}) \in I \times \mathcal{D}_\mathbf{A}$ and all $\mathbf{v} \in \mathcal{D}_\mathbf{A}$, then, from Theorem 3, the trivial solution of the diffusion system is globally asymptotically stable provided that

$$\lim_{t \to \infty} \sup \int_{t_0}^{t} [\lambda \min_i \{\sigma_i(s)\} + \alpha(s)] ds = +\infty. \qquad (76)$$

For the simple case of a single real diffusion equation

$$\partial u / \partial t = \partial^2 u / \partial x^2 + g(u) \qquad (77)$$

defined for $t > 0$ and $\Omega =]0,1[$ with boundary conditions $u(t,0) = u(t,1) = 0$ for all $t \geqslant 0$, we may take $\lambda = \pi^2$. Condition (76) is satisfied when

$$dg(\xi)/d\xi \leqslant \pi^2 \quad \text{for all } \xi \in \mathbb{R}^1. \qquad (78)$$

b. *Symmetric Hyperbolic System*: Let $H = \mathcal{L}_2^n(\Omega)$ as in the diffusion system. Consider a complex hyperbolic system in the form of (60) with

$$\mathbf{A}(t)\mathbf{u} = \sum_{j=1}^{m} \mathbf{A}_{1j}(t,\mathbf{x}) \frac{\partial \mathbf{u}}{\partial x_j}, \qquad (79)$$

where $\mathbf{A}_{1j}(t,\mathbf{x})$, $j = 1, \ldots, m$, are $n \times n$ Hermitian matrices as in (66). Here, we take $\mathcal{D}_\mathbf{A}$ to be the linear subspace of H consisting of continuous functions which are continuously differentiable on $\bar{\Omega}$. For a bounded Ω, we require that they satisfy boundary condition (69). Otherwise, they vanish for sufficiently large values of $\|\mathbf{x}\|_{\mathbb{R}^n}$.

For this system,

$$\text{Re}(\mathbf{Df}(t,\mathbf{u};\mathbf{v}), \mathbf{v}) = -(\mathbf{Q}(t, \circ ;\mathbf{u})\mathbf{v}, \mathbf{v}) \qquad (80)$$

for each $(t,\mathbf{u}) \in I \times \mathcal{D}_\mathbf{A}$ and all $\mathbf{v} \in \mathcal{D}_\mathbf{A}$, where $\mathbf{Q}(t,\mathbf{x};\mathbf{u})$ is given by (74). Similar to the diffusion system, if condition (75) is satisfied with

$$\lim_{t \to \infty} \sup \int_{t_0}^{t} \alpha(s) \, ds = +\infty, \qquad (81)$$

then the trivial solution of the hyperbolic system is globally asymptotically stable.

Finally, for integro-partial differential systems in the form of (60), with $\mathbf{g}(t, \cdot)$ being a nonlinear integral operator having a Gateaux differential on $\mathcal{D}_\mathbf{A}$, we may establish sufficient conditions for global asymptotic stability in the same manner.

ACKNOWLEDGMENTS

[1]Ju. L. Daleckii and M.G. Krein, *Stability of Solutions of Differential Equations in Banach Space*, Trans. Math. Monographs, Vol. 43 (Am. Math. Soc., Providence, R.I., 1974).

[2]G.E. Ladas and V. Lakshmikantham, *Differential Equations in Abstract Spaces* (Academic, New York, 1972).

[3]J.K. Hale, J. Math. Anal. Appl. 26, 39 (1969).

[4]J. Moser (Ed.), *Dynamical System, Theory and Applications*, Lecture Notes in Phys. No. 38 (Springer-Verlag, Berlin, 1975).

[5]G.B. Whitham, *Linear and Nonlinear Waves* (Wiley, New York, 1974).

[6]L. Markus, Rend. Sem. Mat. Politecnico Torino 11, 271 (1952).

[7]L.E. Payne, *Improperly Posed Problems in Partial Differential Equations*, Regional Conf. Series in Appl. Math. No. 22 (Soc. Ind. Appl. Math., Philadelphia, 1975).

[8]P.K.C. Wang, J. Math. Phys. 16, 251 (1975).

[9]M.M. Vainberg, *Variational Methods for the Study of Nonlinear Operators* (Holden-Day, San Francisco, 1964), p. 36.

[10]V. Lakshmikantham and S. Leela, *Differential and Integral Inequalities, Vol. 1* (Academic, New York, 1969), p. 15.

[11]N.N. Krasovskii, Prikl. Mat. Mekh. 18, 735 (1954).

[12]N.N. Krasovskii, Prikl. Mat. Mekh. 21, 309 (1957).

[13]L. Markus and H. Yamabe, Osaka Math. J. 12, 307 (1960).

[14]H. Wilhelmsson, Phys. Rev. A 6, 1973 (1973).

[15]P.K.C. Wang, Nuovo Cimento B 24, 63 (1974).

[16]G.F. Webb, J. Diff. Equations 20, 71 (1976).

[17]L.S. Pontryagin, Izv. Akad. Nauk SSSR, Ser. Mat. 6, 115 (1942).

[18]R. Bellman and K.L. Cooke, *Differential-Difference Equations* (Academic, New York, 1963).

[19]S. Agmon, *Lectures on Elliptic Boundary-Value Problems* (Van Nostrand, New York, 1965).

Relativistic multiple scattering of electromagnetic waves

Svend Berntsen and Günther Johannsen

Institute of Electronic Systems, Aalborg University Center, Denmark
(Received 4 September 1973; final revised manuscript received 2 July 1975)

A general, relativistically exact formulation of scattering of electromagnetic waves by two or more objects in relative, uniform motion is presented. Coupled and separated integral equations are obtained for far-field scattering amplitudes of bodies within a system in terms of known far-field scattering amplitudes of the isolated obstacles. The integral equations are accessible for iteration analogous with the case of nonmoving objects. Explicit expressions and numerical results are given for the two-dimensional problem of two parallel, perfectly conducting cylinders.

I. INTRODUCTION

Scattering of waves by many objects has been intensely investigated in the past.[1] On the other hand, the field of relativistic multiple scattering, i.e., scattering problems involving two or more obstacles which are not at rest in a single inertial frame of reference, is a relatively new subject.

Censor[2] has been concerned with multiple scattering by two cylinders in relative motion using an iterative scheme of successive scattering. The method is based on plane wave integrals representing the scattered fields and leads to a considerable amount of manipulation in order to get explicit expressions. Numerical results are given for a very special case only.

The present work is an attempt to approach this kind of problem in a more general way leading to results which are much more accessible for numerical computation. This is demonstrated in the case of two parallel cylinders of infinite length which are in relative (uniform) motion with respect to each other: As to terms of the first order the amount of computational work is not essentially greater than in the case where both cylinders are at rest. The same case was previously treated by Censor[2] assuming several restrictive conditions (which we do not assume): Both cylinders are thin ($ka \ll 1$) and move on a line which, furthermore, is the direction for the incident plane wave and for the location of the point of observation. Only first order terms in $\beta = v/c$ are taken into account in the evaluation of certain integrals.

Since there is no common rest frame for all scatterers the analysis cannot be based on time-harmonic fields because of the fact that the Lorentz transformation does not in general preserve the property of fields being harmonic in time. This is true even if the scatterers are excited by a plane wave (which is time harmonic in any inertial frame of reference) because the interacting fields are not plane waves.

Our approach is closely related to the fundamental works of Twersky.[3,4] But, even if we hold all scatterers at rest and introduce a three-dimensional notation by suppressing the time coordinate our approach is seen to be different from Twersky's and we believe that the present paper also throws new light on this special case. Our basic physical considerations can shortly be described in this way: Let us fix one of the

scattering objects. It is excited by the incoming signal and the total field due to the presence of the other scatterers.

The latter field can be thought of as being composed by partial fields each "caused by" one of the other objects. These partial fields can in a half-space containing the fixed scatterer be expressed by a superposition of plane waves containing the far-field response of the corresponding scatterer as amplitude. The far-field response of the fixed scatterer corresponding to a single, incident, plane wave is assumed to be known, and, by the principle of superposition, an integral equation is obtained.

Mathematically, our point of departure is the four-dimensional Green's theorem by means of which we express the scattered field as integrals over hypercylinders being parallel to the time axis and representing the scatterers in the four-dimensional space—time manifold. The total field can be written as a sum of partial fields each referring to one scatterer.

By introducing asymptotic expressions for the proper Green's function (both two and three spatial dimensions are treated) far-field "scattering amplitudes" are defined generally.

An exact relationship between the partial fields and their scattering amplitudes are derived by taking the fourfold Fourier transform of the equations expressing the field by means of integrals over hypersurfaces as mentioned before. These relations are used to recast a partial field "coming from" one scatterer and exciting another scatterer as a superposition of plane waves so that the principle of superposition can be applied leading directly to the integral equations. At this point our formalism deviates from that used by Twersky for the case where all scatterers are at rest. Twersky used the Sommerfeld representation of the Hankel function in order to recast the partial fields as a sum of plane waves.

The integral equations which we derive are coupled or separated equations for unknown multiple scattering amplitudes in terms of known (plane wave) scattering amplitudes for the isolated obstacles.

A scalar formalism is in general not possible for electromagnetic, multiple scattering. In order to overcome this problem we make use of the electromagnetic
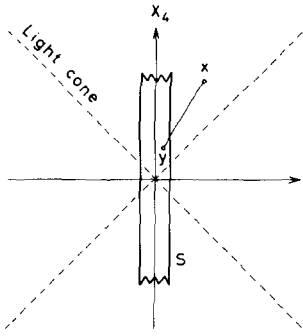
FIG. 1. Location of the scatter in the inertial frame Σ.

potentials which seem to be appropriate for the purpose. In fact, the scalar formulation of the problem can be applied to each component of the 4-potential vector, and the translation into Lorentz-invariant tensor formulation is straightforward.

In the special case, where all scattering objects are at rest in the same frame of reference we indicate how the connection to the results found by Twersky (who did not use the electromagnetic potentials, cf. Ref. 7) can be established. Finally, we indicate how the formalism can be extended to an arbitrary number of obstacles moving with different velocities.

II. ONE SCATTERER

Consider a scattering object excited by an arbitrary signal and an inertial frame of reference Σ, where the object is at rest. See Fig. 1. The coordinates of a world point in Σ are denoted by $x = (x_1, x_2, x_3, x_4)$ $= (\bar{x}, x_4)$, where $\bar{x} = (x_1, x_2, x_3)$ are spatial Cartesian coordinates. The time coordinate x_4 is defined by $x_4 = ict$, where t is the time and c the velocity of light in a vacuum. The four-dimensional scalar product of two 4-vectors p and q is denoted by $p \cdot q = \bar{p} \cdot \bar{q} + p_4 q_4$.

In Σ the scatterer is represented by a cylindrical hypersurface parallel to the x_4 axis.

Applying the four-dimensional Green's theorem the scattered field $u(x)$ satisfying the homogeneous wave equation (representing a Cartesian component of the electromagnetic field or a component of the 4-vector potential, cf. Sec. III) can be expressed by an integral over the hypersurface of the object,

$$u(x) = i \int_{-i\infty}^{i\infty} dy_4 \oint_S dS(\bar{y})[h(x-y)\partial_n u(y) - u(y)\partial_n h(x-y)].$$
(1)

In Eq. (1) S is identical to the spatial surface of the scattering obstacle (or any surface enclosing it). Because the hypercylinder is parallel to the x_4 axis, ∂_n denotes the spatial directional derivative in the direction of the outward normal of S. Finally, h is the retarded Green's function given by[5]

$$h(x) = \frac{\theta(-ix_4)}{4\pi} \frac{\delta(|\bar{x}| + ix_4)}{|\bar{x}|} = \frac{\theta(-ix_4)}{2\pi} \delta(x \cdot x),$$
(2)

where $\theta(-ix_4) = 1$ for $-ix_4 \geqslant 0$ and $= 0$ elsewhere, and

$$\left(\nabla^2 + \frac{\partial^2}{\partial x_4^2}\right) h(x) = -\delta(x) = -\delta(\bar{x})\delta(-ix_4).$$
(3)

The Fourier transform of h is defined by

$$\tilde{h}(p) = -i \int d^4x h(x) e^{-ip \cdot x} = \frac{1}{p \cdot p - \epsilon p_4}, \quad (\epsilon \to 0^+),$$
(4)

which assumes the causal nature of the Green function. (We use tildes to denote Fourier transforms).

Taking the fourfold Fourier transform of Eq. (1), we obtain

$$\tilde{u}(p) = \tilde{h}(p)G(p),$$
(5)

where G is defined by

$$G(p) = -\oint_S dS(\bar{y})[e^{-i\bar{p}\cdot\bar{y}}\partial_n\hat{u}(\bar{y}, p_4) - \hat{u}(\bar{y}, p_4)\partial_n e^{-i\bar{p}\cdot\bar{y}}],$$
(6)

and where the Fourier transform with respect to the variable y_4 is given by

$$\hat{u}(\bar{y}, p_4) = -i \int dy_4 u(y) e^{-ip_4 y_4}.$$
(7)

As to the second term under the integral sign in (6) integration by parts has been used.

For $p \cdot p = 0$, $G(p)$ can be given a physical interpretation. In order to do this we will look at a definite frequency. Taking the Fourier transform of Eq. (1) with respect to the variable x_4 we find

$$\hat{u}(\bar{x}, p_4) = -\oint_S dS(\bar{y})[\hat{h}(\bar{x} - \bar{y}, p_4)\partial_n\hat{u}(\bar{y}, p_4)$$
$$- \hat{u}(\bar{y}, p_4)\partial_n\hat{h}(\bar{x} - \bar{y}, p_4)],$$
(8)

where

$$\hat{h}(\bar{x}, p_4) = \frac{1}{4\pi} \frac{e^{p_4|\bar{x}|}}{|\bar{x}|}.$$
(9)

The relation (8) might be obtained directly by means of the three-dimensional Green's theorem.

Consider the case, where $|p_4||\bar{x} - \bar{y}| \gg 1$. Defining a vector $\bar{p} = -ip_4\bar{x}/|\bar{x}|$ (which implies $p \cdot p = 0$) we have approximately $-ip_4|\bar{x} - \bar{y}| \simeq -ip_4|\bar{x}| - \bar{p}\cdot\bar{y}$.

From this and (8) we obtain

$$\hat{u}(\bar{x}, p_4) \simeq -\frac{e^{p_4|\bar{x}|}}{4\pi} \oint_S dS(\bar{y})\left(\frac{e^{-i\bar{p}\cdot\bar{y}}}{|\bar{x} - \bar{y}|} \partial_n\hat{u}(\bar{y}, p_4)\right.$$
$$\left. - \hat{u}(\bar{y}, p_4)\partial_n \frac{e^{-i\bar{p}\cdot\bar{y}}}{|\bar{x} - \bar{y}|}\right).$$
(10)

In the case of three spatial dimensions we are interested in the situation where, futhermore, $|\bar{x}| \gg |\bar{y}|$ for all y on the surface S of the scatterer. We find that

$$\hat{u}(\bar{x}, p_4) \simeq \hat{h}(\bar{x}, p_4)G(\bar{p}, p_4),$$
(11)

where $G(\bar{p}, p_4)$ is the general "scattering amplitude" expressing how a single frequency component (with frequency $-icp_4$) of the scattered field for large distance depends on the direction as given by $\bar{p}$. Notationally we have suppressed the dependence of G on the exciting field which is not necessarily a plane wave.

In the case of two spatial dimensions the demand $|\bar{x}| \gg |\bar{y}|$ for "all" $\bar{y}$ has no physical meaning. Therefore, we go back to (8) and integrate with respect to one of the spatial coordinates, y_1, for example. Obviously, we can assume that u is independent of y_1, and we will use the notation $u(x, k_4)$ with $x = (x_2, x_3)$. Furthermore, S can be assumed to be of infinite extension in the direction of the x_1 axis.
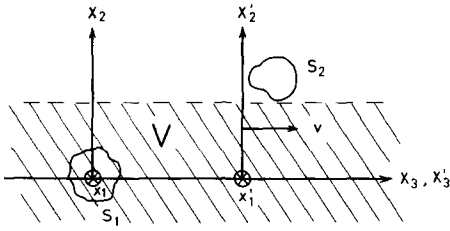
FIG. 2. Location of two moving scatterers in the inertial frames Σ and Σ'.

Using the well-known representation of zero order Hankel functions of the first and second kind as given by

$$\int_{-\infty}^{\infty} \frac{\exp(p_4|\bar{x} - \bar{y}|)}{|\bar{x} - \bar{y}|} dy_1 = i\pi H_0^{(\nu)}(|p_4| \ |\underline{x} - \underline{y}|),$$

$$\nu = \begin{cases} 1 & \text{for } p_4/i > 0, \\ 2 & \text{for } p_4/i < 0, \end{cases} \tag{12}$$

Eq. (8) reduces to an expression for u which also can be obtained by means of the well-known two-dimensional Green's theorem,

$$\hat{u}(\underline{x}, p_4) = -(i/4)\oint_S dS(\underline{y})[H_0^{(\nu)}(|p_4| \ |\underline{x} - \underline{y}|)\partial_n\hat{u}(\underline{y}, p_4) - \hat{u}(\underline{y}, p_4)$$
$$\cdot \partial_n H_0^{(\nu)}(|p_4| \ |\underline{x} - \underline{y}|)]. \tag{13}$$

If $|p_4| \ |\underline{x} - \underline{y}| \gg 1$ and $|x| \gg |y|$ for all y on the scatterer S (a two-dimensional closed curve), we can introduce asymptotic expressions for the Hankel functions which leads to

$$\hat{u}(\underline{x}, p_4) \simeq \frac{1}{2\sqrt{2\pi}} \frac{\exp[\pm i(|p_4| \ |x| + \pi/4)]}{(|p_4| \ |x|)^{1/2}} G(p, p_4),$$

$$+\text{for } p_4/i > 0, \quad -\text{for } p_4/i < 0. \tag{14}$$

Let $u(x;k)$ denote the scattered field corresponding to an incoming, plane wave given by $\exp(ik \cdot x)$, where $k \cdot k = 0$ as a consequence of the fact that the plane wave obeys the homogeneous wave equation. Since $-ik_4$ is assumed to be real, the components of $\bar{k}$ may be complex provided $k \cdot k = 0$ is satisfied.

$u(x;k)$ is harmonic in time so, we may write

$$u(x;k) = w(\bar{x};k) \exp(ik_4 x_4) \tag{15}$$

and

$$\hat{w}(\bar{x}, p_4;k) = 2\pi\delta(ip_4 - ik_4)w(\bar{x};k). \tag{16}$$

If the general scattering amplitude for the field (15) is denoted by $G(p;k)$ we get from (6) and (16) that

$$G(p;k) = 2\pi\delta(ip_4 - ik_4)g(\bar{p};k), \tag{17}$$

where we have introduced the plane wave scattering amplitude

$$g(\bar{p};k) = -\oint_S dS(\bar{y})[e^{-i\bar{p}\cdot\bar{y}}\partial_n w(\bar{y};k) - w(\bar{y};k)\partial_n e^{-i\bar{p}\cdot\bar{y}}]. \tag{18}$$

The concept of scattering amplitude as given by (18) is identical with that defined by Twersky.[3,4]

III. TWO SCATTERERS

When we are dealing with two or more scatterers in relative motion to each other we have to face the fact

that a scalar formalism as used in Sec. II is no longer sufficient. Only for a very special configuration which will be treated in the next section can a scalar formalism be applied.

Four-dimensional tensor notation will be used. Greek subscripts run from 1 to 4 and repeated subscripts obey the summation convention. Commas in these subscripts denote partial differentiation with respect to the coordinates (or covariant differentiation since the metric tensor is independent of the coordinates). The metric tensor is given by the Kronecker symbol $\delta_{\lambda\mu}$ (Cartesian spatial coordinates are used) and therefore, we do not distinguish between contravariant and covariant tensors.

Consider two scattering objects S_1 and S_2 which are at rest in the inertial frames of reference denoted by Σ and Σ', respectively. Without loss of generality we may choose the Lorentz transformation to be given by

$$x_\lambda = a_{\lambda\mu}x'_\mu, \tag{19}$$

where

$$a_{\lambda\mu} = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & \gamma & -i\gamma\beta \\ 0 & 0 & i\gamma\beta & \gamma \end{pmatrix}, \tag{20}$$

$\beta = v/c$, $\gamma = (1 - \beta^2)^{-1/2}$, and v is the relative velocity between the scatterers, cf. Fig. 2.

The situation can also be illustrated in a four-dimensional space—time manifold. The scatterers are then represented by two hypercylinders each parallel to its time axis and displaced in the direction of the $x_2 - x'_2$ axis, cf. Fig. 3.

We want to make use of the electromagnetic potentials $\bar{A}$, φ in order to elaborate the general formalism.

It is well-known that the equations
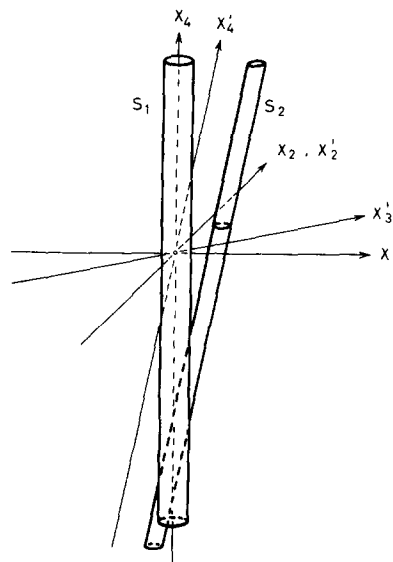
FIG. 3. Location of two scatterers in four-dimensional space—time manifold.

1423   J. Math. Phys., Vol. 17, No. 8, August 1976

S. Berntsen and G. Johannsen   1423

$$\vec{B} = \nabla \cdot \vec{A}, \quad \vec{E} = -\nabla\varphi - \frac{\partial \vec{A}}{\partial t} \tag{21}$$

as well as the Lorentz convention

$$\nabla \cdot \vec{A} + \frac{1}{c^2}\frac{\partial \varphi}{\partial t} = 0 \tag{22}$$

are Lorentz invariant. Furthermore, $u_\lambda = (\vec{A}, i\varphi/c)$ is a tensor of the first order (4-vector) called the 4-potential.

In tensor notation (22) is given by

$$u_{\lambda,\lambda} = 0, \tag{23}$$

and (21) can be written as

$$F_{\lambda\mu} = c(u_{\lambda,\mu} - u_{\mu,\lambda}). \tag{24}$$

The components of the field tensor $F_{\lambda\mu}$ may be found in Ref. 6.

Each component of the 4-potential satisfies at any point outside the scatterers, in any inertial frame of reference the homogeneous wave equation, i.e.,

$$u_{\lambda,\mu\mu} = 0. \tag{25}$$

Applying the four-dimensional Green's theorem the 4-potential of the total scattered field u_λ can be written as a sum

$$u_\lambda(x;\Psi) = u_\lambda^{(1)}(x;\Psi) + u_\lambda^{(2)}(x;\Psi), \tag{26}$$

where $u_\lambda^{(1)}$ and $u_\lambda^{(2)}$ are 4-potential fields given by integrals over the hypercylinders of the objects. In the respective rest frame Σ and Σ' these integrals are quite analogous with (1) involving the surfaces S_1 and S_2 respectively.

In (26) Ψ denotes the dependence on the incoming, exciting field $\Psi_\lambda(x)$. The field $u_\lambda^{(1)}$ can be interpreted as the response from scatterer S_1, due to the incoming field Ψ_λ plus the "multiple scattered response" due to the field $u_\lambda^{(2)}$.

The formalism which was developed in Sec. II can be applied to each component of the fields $u_\lambda^{(1)}$ and $u_\lambda^{(2)}$. We will do it in order to work out the integral equations for these unknown fields or for the more suitable functions $G_\lambda^{(1)}$ and $G_\lambda^{(2)}$ defined by means of (6) and representing the multiple scattering amplitudes.

For this purpose consider the situation at scatterer S_1, for example. In part V (hatched in Fig. 2) of the space (which contains S_1) we can represent the total incoming field on S_1 as superposition of plane waves (obeying the wave equation). In order to do this we consider the fourfold Fourier transform of $u_\lambda^{(2)}$ given by (5),

$$u_\lambda^{(2)}(x;\Psi) = -i\left(\frac{1}{2\pi}\right)^4 \int d^4p\, \tilde{h}(p) G_\lambda^{(2)}(p;\Psi) \exp(ip_\mu x_\mu). \tag{27}$$

The Green's function defined by (2) is invariant (and covariant), i.e., $h'(x') = h(x)$ $[= h'(x)]$ and so is the Fourier transform $\tilde{h}$. Furthermore, $u_\lambda^{(2)}$ is a 4-vector implying that $G_\lambda^{(2)}$ is a 4-vector, and by means of (23) we derive from (27) that

$$p_\lambda G_\lambda^{(2)}(p, \Psi) = 0. \tag{28}$$

In the rest frame Σ' of scatterer S_2, $G_\lambda'^{(2)}(p';\Psi)$ is defined by means of (6) and is analytical in $p_2' = p_2$. This

implies that $G_\lambda^{(2)}(p;\Psi)$ is also analytical in p_2 and therefore, for all $\vec{x}$ in V we can close the contour of p_2 integration in (27) in the lower half of the complex p_2 plane because we can assume that $x_2 - y_2 < 0$. (It means that the scatterers do not come into collision at any time, cf. Fig. 2.)

Denoting the closed contour in the lower half of the complex p_2 plane by $C-$ the field $u_\lambda^{(2)}$ can for all x in V be written as a superposition of plane waves satisfying the homogeneous wave equation,

$$u_\lambda^{(2)}(x;\Psi) = -i\left(\frac{1}{2\pi}\right)^4 \int_{C-} d^4p\, G_\lambda^{(2)}(p;\Psi)\tilde{h}(p)\exp(ip_\mu x_\mu). \tag{29}$$

When $|p_4|^2 > (p_1)^2 + (p_3)^2$, the plane waves in (29) are real, i.e., nonattenuated in all coordinates. When $|p_4|^2 < (p_1)^2 + (p_3)^2$ the field components of (29) are inhomogeneous plane waves, propagating in a direction orthogonal to the x_2 axis and decreasing exponentially in the direction of the negative x_2 axis.

Equation (28) shows that the plane waves in (29) obey the Lorentz condition so that they can be interpreted as 4-potentials. This is in agreement with the physical interpretation of $G_\lambda^{(2)}$ as far-field scattering amplitudes of the 4-potential $u_\lambda^{(2)}$, cf. Sec. II.

Before integral equations can be set up by means of the principle of superposition it is necessary to define responses and scattering amplitudes for the isolated obstacles as we did in Sec. II, Eqs. (15)—(18). But, in the present case the situation is a little more complicated because we are dealing with vector fields.

Consider a plane 4-potential wave

$$A_\lambda \exp(iq_\mu x_\mu) \tag{30}$$

satisfying the homogeneous wave equation, i.e.,

$$q_\lambda q_\lambda = 0 \tag{31}$$

and the Lorentz convention (23), i.e.,

$$A_\lambda q_\lambda = 0. \tag{32}$$

A response to (30) depends not only on q_λ but also on the 4-vector A_λ.

Because of (32) the set of possible amplitudes A_λ forms a three-dimensional space. Furthermore, $A_\lambda = q_\lambda$ is a possible amplitude 4-vector leading to the noninteresting case, where all components of the electromagnetic field vanish, which follows from Eq. (24). Therefore, the set of possible amplitudes in (30) which are of interest forms a two-dimensional vector space.

In the light of these remarks we define for given q_λ two basic amplitude 4-vectors $b_\lambda^{(\alpha)}(q)$, $\alpha \in \{1,2\}$ satisfying (32) and the additional conditions

$$b_\lambda^{(\alpha)} b_\lambda^{(\alpha)} = 1, \quad b_\lambda^{(1)} b_\lambda^{(2)} = 0. \tag{33}$$

[This is always possible which may be seen in this way: Consider an arbitrary frame of reference Σ^*, choose $b_4^{(\alpha)*} = 0$ and choose the spatial parts of the 4-vectors in question as two unit vectors perpendicular to each other and to $\vec{q}^*$. Then (32) and (33) are satisfied in Σ^* and therefore in any system of reference Σ because we are dealing with tensor equations.]

As mentioned before we have to introduce some response functions referring to the isolated obstacles. Let an isolated scatterer, S_1 for example, be excited by the incident field Ψ_λ. The scattered field is denoted by $u_{0\lambda}^{(1)}(x;\Psi)$ and the corresponding general scattering amplitude by $G_{0\lambda}^{(1)}(p;\Psi)$.

In the special case, where Ψ_λ is a basic plane wave $b_\lambda^{(\kappa)}(q)\exp(iq_\mu x_\mu)$ the scattered field is denoted by $u_{0\lambda}^{(1)}(x;q,\kappa)$ and the corresponding genera scattering amplitude by $G_{0\lambda}^{(1)}(p;q,\kappa)$. According to (17) we have

$$G_{0\lambda}^{(1)}(p;q,\kappa) = 2\pi\delta(ip_4 - iq_4)g_{0\lambda}^{(1)}(\vec{p};q,\kappa).\qquad(34)$$

It is observed that this equation only can be used in the rest frame Σ of scatterer S_1, the plane wave scattering amplitude $g_{0\lambda}^{(1)}$ is not a 4-vector.

The field $u_\lambda^{(2)}$ is a superposition of plane waves as given by (29) satisfying the Lorentz convention (28). Each amplitude 4-vector $\tilde{h}(p)G_\lambda^{(2)}(p;\Psi)$ of these plane waves can be decomposed into two components proportional to the introduced basic amplitude 4-vectors $b_\mu^{(\kappa)}(p)$. Hence, by the principle of superposition the response from scatterer S_1 to $u_\lambda^{(2)}$ plus the incoming field Ψ_λ is found to be

$$u_\lambda^{(1)}(x;\Psi) = u_{0\lambda}^{(1)}(x;\Psi) - i\left(\frac{1}{2\pi}\right)^4 \int_{C_-} d^4p\tilde{h}(p)$$

$$\times \sum_{\kappa=1}^2 u_{0\lambda}^{(1)}(x;p,\kappa)G_\mu^{(2)}(p;\Psi)b_\mu^{(\kappa)}(p).\qquad(35)$$

The corresponding equation expressing the response from scatterer S_2 might be found in the same way.

Taking the fourfold Fourier transform of Eq. (35) we derive by means of (5) the integral equation

$$G_\lambda^{(1)}(k;\Psi) = G_{0\lambda}^{(1)}(k;\Psi) - i\left(\frac{1}{2\pi}\right)^4 \int_{C_-} d^4p\tilde{h}(p)\sum_{\kappa=1}^2 G_{0\lambda}^{(1)}(k;p,\kappa)$$

$$\times G_\mu^{(2)}(p;\Psi)b_\mu^{(\kappa)}(p).\qquad(36)$$

In the same way we derive the integral equation

$$G_\lambda^{(2)}(k;\Psi) = G_{0\lambda}^{(2)}(k;\Psi) - i\left(\frac{1}{2\pi}\right)^4 \int_{C_+} d^4p\tilde{h}(p)$$

$$\times \sum_{\kappa=1}^2 G_{0\lambda}^{(2)}(k;p,\kappa)\, G_\mu^{(1)}(p;\Psi)b_\mu^{(\kappa)}(p),\qquad(37)$$

where C_+ is the closed contour in the upper half of the complex p_2 plane.

Equations (36) and (37) are coupled integral equations for the unknown functions $G_\lambda^{(1)}$ and $G_\lambda^{(2)}$, which in principle can be calculated by means of these integral equations if the plane wave scattering amplitudes as well as the general scattering amplitudes referring to the isolated obstacles are known.

It is not difficult to separate the coupled integral equations with respect to the obstacles. If the zero order term of (37) is inserted into (36), we obtain a first order contribution to the scattered field given by

$$G_{I\lambda}^{(1)}(k;\Psi) = - i\left(\frac{1}{2\pi}\right)^4 \int_{C_-} d^4p\tilde{h}(p)\sum_{\kappa=1}^2 G_{0\lambda}^{(1)}(k;p,\kappa)$$

$$\times G_{0\mu}^{(2)}(p;\Psi)b_\mu^{(\kappa)}(p).\qquad(38)$$

Eliminating $G_\mu^{(2)}$ in (36) leads to

$$G_\lambda^{(1)}(k;\Psi) = G_{0\lambda}^{(1)}(k;\Psi) + G_{I\lambda}^{(1)}(k;\Psi) - i\left(\frac{1}{2\pi}\right)^4 \int_{C_+} d^4p\tilde{h}(p)$$

$$\times \sum_{\kappa=1}^2 b_\mu^{(\kappa)}(p)G_{I\lambda}^{(1)}(k;p,\kappa)G_\mu^{(1)}(p;\Psi).\qquad(39)$$

The analogous equations referring to the other scatterer are found in the same way. Equations (36)—(39) are tensor equations, i.e., they do not change their form under Lorentz transformations as given by (19).

The scattered electromagnetic far-field is easily derived from the general scattering amplitudes by means of (24): Let an observer be situated in an arbitrary inertial frame of reference which not necessarily is one of the rest frames of the scatters. Let S_σ, $\sigma \in \{1,2\}$ be a closed surface enclosing one of the scattering objects at any time. Strictly speaking, such a closed surface exists only if the object is at rest in the observers frame of reference but, the main contribution to the field may arise from a limited time interval. This is the case for multiple scattered fields by two moving objects, where the main contribution can be expected to arise from a time interval, where the scattering objects are close together.

In the far-field region the scattered 4-potential related to S_σ can be found by (9)—(11) [or (14) in the case of two spatial dimensions]. We derive that

$$u_\lambda^{(\sigma)}(x;\Psi) \simeq f(r)\int_{-\infty}^{\infty} d(k_4/i)G_\lambda^{(\sigma)}(k;\Psi)e^{ik_\mu x_\mu}, \quad \sigma \in \{1,2\},\qquad(40)$$

where $f(r) = 1/4\pi r$, $r = |\vec{x}|$, $\vec{k} = (\vec{x}/|\vec{x}|)k_4/i$ [or $f(r) = \exp(\pm i\pi/4)/(8\pi|k_4|r)^{1/2}$, + for $k_4/i > 0$, − for $k_4/i < 0$, $r = |x|$, $\vec{k} = (x/|x|)\cdot k_4/i$ in the case of only two spatial dimensions].

As mentioned in Sec. II, formula (40) preassumes that the farfield conditions $|k_4|\,|\vec{x} - \vec{y}| \gg 1$ and $|\vec{x}| \gg |\vec{y}|$ for all $\vec{y}$ on S_σ are satisfied for all k_4/i which are essential for the integral in (40). From (40) and (24) the electromagnetic far-field components corresponding to S_σ are obtained immediately:

$$F_{\lambda\mu}^{(\sigma)}(x;\Psi) \simeq icf(r)\int_{-\infty}^{\infty} d(k_4/i)[k_\mu G_\lambda^{(\sigma)}(k;\Psi)$$

$$- k_\lambda G_\mu^{(\sigma)}(k;\Psi)]e^{ik_\nu x_\nu}, \sigma \in \{1,2\}.\qquad(41)$$

For numerical purposes we will introduce plane wave scattering amplitudes into (38) by means of (34) and its analog referring to scatterer S_2. We also assume that the exciting field Ψ_λ is a basic plane wave with amplitude 4-vector $b_\lambda^{(\iota)}(q)$. Then,

$$G_{I\lambda}^{(1)}(k;q,\iota) = - i\left(\frac{1}{2\pi}\right)^2 \int_{C_-} d^4p\tilde{h}(p)\delta(ip_4 - ik_4)\delta(ip_4' - iq_4')$$

$$\times \sum_{\kappa=1}^2 g_{0\lambda}^{(1)}(\vec{k};p,\kappa)\sum_{\mu=1}^4 g_{0\lambda}'^{(2)}(\vec{p}';q',\iota)b_\mu'^{(\kappa)}(p'),\qquad(42)$$

where $\iota \in \{1,2\}$.

Equations (40)—(42) are not covariant equations. (42) refers to the rest frames of the scatterers, and the approximations in (40) and (41) depend on the frame of reference.

1425   J. Math. Phys., Vol. 17, No. 8, August 1976

S. Berntsen and G. Johannsen   1425

The first ("zero order") term on the right side of (36) or of (37) is the response of the exciting field corresponding to isolated obstacles.

It is observed that the functions $G_\mu^{(1)}$ and $G_\mu^{(2)}$ on the right sides of the integral equations (36) and (37) are restricted to the "cone" $p_\lambda p_\lambda = 0$, where p_λ may be complex. Furthermore, it is seen from (11) that as far as we are concerned with scattered far-fields we are not interest in the functions $G_\lambda^{(1)}(k;\Psi)$ and $G_\lambda^{(2)}(k;\Psi)$ for all values of k_λ but only for real values of $\bar{k}$ and k_4/i on the cone $k_\lambda k_\lambda = 0$.

It is of practical interest to know how the fields μ_λ transform under a space—time translation as given by

$$x_\lambda^t = x_\lambda - d_\lambda. \tag{43}$$

If $u_\lambda(x;q,\imath)$ is any field response due to an exciting field $b_\lambda^{(1)}(q) \cdot \exp(iq_\mu x_\mu)$, and $u_\lambda^t(x^t;q,\imath)$ denotes the corresponding field response in the translated frame of reference due to the excitation $b_\lambda^{(1)}(q) \cdot \exp(iq_\mu x_\mu^t)$ $= b_\lambda^{(1)}(q)\exp(iq_\mu x_\mu - iq_\mu d_\mu)$ we conclude by means of linearity

$$u_\lambda^t(x^t;q,\imath) = \exp(-iq_\mu d_\mu)u_\lambda(x;q,\imath). \tag{44}$$

A fourfold Fourier transformation of this equation leads to

$$G_\lambda^t(p;q,\kappa) = \exp[-i(q_\mu - p_\mu)d_\mu]G_\lambda(p;q,\kappa) \tag{45}$$

for the corresponding scattering amplitudes as defined by (5).

It is readily seen that the formalism can be extended to an arbitrary number of scattering objects. If the objects are not moving in the same direction spatial rotations have to be taken into account, of course.

IV. TWO-DIMENSIONAL, SCALAR SCATTERING

Consider the case, where we are concerned with only two spatial coordinates x_2 and x_3, for example, corresponding to a configuration, where the scattering obstacles are cylinders of infinite length parallel with the x_1 axis, cf. Fig. 2. The formalism developed in the foregoing sections can readily be adjusted to this case.

Inspection of formula (6) shows that the function G now involves a factor $2\pi\delta(p_1)$ the remainder being independent of p_1. From this we conclude that the integral equations (36) and (37) take the form

$$G_\lambda^{(1)}(k;\Psi) = G_{0\lambda}^{(1)}(k;\Psi) - i\left(\frac{1}{2\pi}\right)^3 \int d^3p\tilde{h}(p)$$

$$\times \sum_{\kappa=1}^{2} G_{0\lambda}^{(1)}(k;p,\kappa)G_\mu^{(2)}(p;\Psi)b_\mu^{(\kappa)}(p), \tag{46}$$

$$G_\lambda^{(2)}(k;\Psi) = G_{0\lambda}^{(2)}(k;\Psi) - i\left(\frac{1}{2\pi}\right)^3 \int d^3p\tilde{h}(p)$$

$$\times \sum_{\kappa=1}^{2} G_{0\lambda}^{(2)}(k;p,\kappa)G_\mu^{(1)}(p;\Psi)b_\mu^{(\kappa)}(p), \tag{47}$$

where $d^3p = dp_2dp_3dp_4$, the G functions and $b^{(\kappa)}$ are independent of p_1 and $\tilde{h}$ is still given by (5) with $p_1 = 0$. Other equations in Sec. III can readily be adjusted in the same way.

Next, we will look at a special case which is particularly simple because only one component of the 4-potential is involved, i.e., the problem is scalarized.

Let the exciting field Ψ_λ be a plane wave perpendicular to and polarized parallel with the cylinders,

$$\Psi_\lambda = \begin{pmatrix} 1 \\ 0 \\ 0 \\ 0 \end{pmatrix} \exp(iq_\mu x_\mu), \quad q_1 = 0 \tag{48}$$

The Lorentz convention (32) is satisfied, and (48) holds in any inertial frame of reference.

We may assume that the scattered field from an isolated cylinder consists of a single component parallel with the cylinder, i.e., $G_{0\lambda}^{(\sigma)} = 0$ for $\sigma \in \{1,2\}$, $\lambda \in \{2,3,4\}$. Imagine now that the integral equations (46) and (47) are solved by iteration. By induction we conclude that also $G_\lambda^{(\sigma)} = 0$ for $\sigma \in \{1,2\}$, $\lambda \in \{2,3,4\}$. This shows that only the first component of the potentials are involved and therefore, we erase the coordinate subscript from now on, i.e., we write $G^{(\sigma)}$ instead of $G_1^{(\sigma)}$, and $G_I^{(\sigma)}$ instead of $G_{I1}^{(\sigma)}$, etc.

If the Ith order contribution to the scattered field as given by (42) is adjusted to the present situation we obtain

$$G_I^{(1)}(k;q) = -(i/2\pi)\int_{C_-} d^3p\tilde{h}(p)\delta(ip_4 - ik_4)\delta(ip_4' - iq_4')$$

$$\times g_0^{(1)}(k;p)g_0^{t(2)}(p';q')\exp[i(q_2 - p_2)d] \tag{49}$$

and in quite the same way, corresponding to S_2,

$$G_I'^{(2)}(k';q') = -\frac{i}{2\pi}\int_{C_+} d^3p'\tilde{h}(p')\delta(ip_4' - ik_4')\delta(ip_4 - iq_4)$$

$$\times g_0^{t(2)}(k';p')g_0^{(1)}(p;q)\exp[i(p_2 - k_2)d]. \tag{50}$$

The phase factors $\exp[i(q_2 - p_2)d]$ in (49) and $\exp[i(p_2 - k_2)d]$ in (50) are introduced by means of a coordinate translation (43) applied to $g_0'^{(2)}$ so that the plane wave scattering amplitude $g_0^{t(2)}$ refers to a rest frame for S_2 which spatial origin passes that of the rest frame for S_1 in a distance d (at the time $x_4/ic = x_4'/ic = 0$).

Though $g_0^{(1)}$ and $g_0^{t(2)}$ and the form of the right sides in (49) and (50) refer to inertial frames of reference, where the scattering cylinders are at rest respectively, both equations can easily be transformed to an arbitrary system of reference. This is so because the quantities on the left sides are known to be invariants. In particular we have $G_I'^{(2)}(k';q') = G_I^{(2)}(k;q)$.

The first-order terms (49) and (50) are readily accessible for numerical computation ($g_0^{(1)}$ and $g_2^{t(2)}$ must be available, of course). Making use of the δ functions and the pole in $\tilde{h}(p)$ all integrations $d^3p = dp_2dp_3dp_4$ can be carried out. It turns out that

$$G_I^{(1)}(k;q) = -\frac{i\gamma}{2\beta p_{2-}}g_0^{(1)}(k;p_-)g_0^{t(2)}(p_-;q')$$

$$\times \exp[i(q_2 - p_{2-})d], \tag{51}$$

$$G_I'^{(2)}(k';q') = \frac{i}{2\beta p_{2+}'}g_0^{t(2)}(k';p_+')g_0^{(1)}(p_+;q)$$

$$\times \exp[-i(k_2 - p_{2+})d], \tag{52}$$
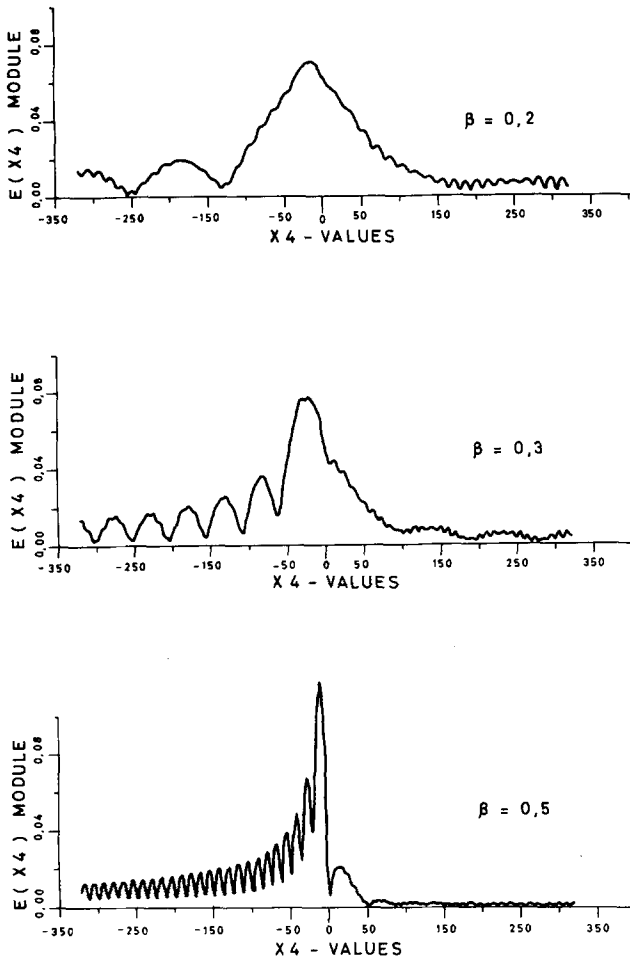
S. Berntsen and G. Johannsen   1426

FIG. 4. Far-field amplitude of the multiple scattered field of the first order for two parallel, perfectly conducting, circular cylinders of infinite length. One of the cylinders is at rest in the frame of observation and the other cylinder moves perpendicular to the cylinders and to the direction in which an observer is situated. Furthermore, the incident plane wave comes from the direction of observation $[x_1 = x_3 = q_1 = q_3 = 0$, $q_2 = q_4/i = 1$, i.e., "back scattering," cf. (48) and Fig. 2]. The cylinders have equal radius $a = \lambda/\pi$ (λ is the wavelength of the incident field), and the shortest distance between the center-lines of the cylinder is $d = 4\lambda/\pi$. During the normalized time $x_4 = |q_4| ct$ (observers frame) the moving object traverses a distance given by $\lambda \beta x_4/2\pi$, where $\beta = v/c$ is the normalized velocity. At the time $x_4 = 0$ (retarded with respect to the position of the observer) the two objects are nearest to each other.

where $p_{3-} = (k_4 - q_4'/\lambda)/(i\beta)$, $p_{3+} = (q_4 - k_4'/\lambda)/(i\beta)$, $p_{2-} = -(-p_{3-}^2 - k_4^2)^{1/2}$, and $p_{2+} = (-p_{3+}^2 - q_4^2)^{1/2}$. Furthermore, it is assumed that $q_4/i > 0$.

In order to find the electromagnetic far-field we have to insert (51) and (52) into (41) and add the results in a proper way. Let us find the electric field vector given by[6] $E_1 = F_{41}/i$, $E_2 = E_3 = 0$. Remembering that $G_{I\lambda}^{(\sigma)} = 0$ for $\lambda \in \{2, 3, 4\}$, $\sigma \in \{1, 2\}$ and that $G_{I1}^{(\sigma)}$ is denoted by $G_I^{(\sigma)}$, etc., we derive for the multiple scattered far-field of first order

$$E_I^{(\sigma)}(x;q) \simeq c \, \frac{e^{\pm i\pi/4}}{\sqrt{8\pi}} \int_{-\infty}^{\infty} d(k_4/i) \, \frac{e^{ik_\mu x_\mu}}{(|k_4| \, |x|)^{1/2}} \, (k_4/i) G_I^{(\sigma)}(k;q), \tag{53}$$

where $\sigma \in \{1, 2\}$, $+$ for $k_4/i > 0$ and $-$ for $k_4/i < 0$.

Equation (53) is valid in any inertial frame of reference, where the far-field conditions as mentioned in connection with Eq. (40) are fulfilled.

If an observer is placed such that (53) is a good approximation to the far-field for both values of σ, it follows that the first order term of the total multiple scattered far-field is given by

$$E_I(x;q) \simeq c \, \frac{e^{\pm i\pi/4}}{\sqrt{8\pi}} \int_{-\infty}^{\infty} d(k_4/i)$$
$$\times \frac{\exp[i(|k| \, |x| - x_4/i)k_4/i]}{(|k_4| \, |x|)^{1/2}} \, (k_4/i)[G_I^{(1)}(k;q) + G_I^{(2)}(k;q)], \tag{54}$$

where we have made use of $k = (x/|x|)(k_4/i)$, cf. Eq. (40).

Another approximation to the far-field is obtained by using (53) in the rest frame of reference of the respective scatterer and then transforming both contributions to the observers frame. We will do this for the case where the observer does not move relative to S_1 so that $E_I'^{(2)}$ is to be transformed.

Using (20) and (24) we derive that $E^{(2)}(x;q) = F_{41}/i$
$= a_{4\mu} a_{1\nu} F_{\mu\nu}'/i = a_{4\mu} F_{\mu1}'/i = (a_{43} F_{31}' + a_{44} F_{41}')/i = \gamma(\beta F_{31}' + F_{41}'/i)$. From this, Eq. (41) and the relation $k' = (x'/|x'|)k_4'/i$ we obtain

$$E_I^{(2)}(x;q) \simeq c \, \frac{e^{\pm i\pi/4}}{\sqrt{8\pi}} \int_{-\infty}^{\infty} d(k_4'/i) \, \frac{\exp[i(|k'| \, |x'| - x_4'/i)]}{(|k_4'| \, |x'|)^{1/2}}$$
$$\times (k_4'/i)\left(1 + \beta \, \frac{k_3'}{k_4'/i}\right) G_1'^{(2)}(k';q') \tag{55}$$

which is to be added to (53) with $\sigma = 1$ in order to find the first order term of the total, multiple scattered far-field. [It can be shown that Eq. (54) or (55) contains Eq. (17) of Ref. 2 as a special case.]

In Fig. 4 numerical results are given based on formula (54), they expose an interesting feature: At very high velocities such as $\beta = 0.5$ the interaction between the two objects decreases rapidly when the moving object has passed the point of shortest distance between the scatterers.

It is worthwhile to remark that as far as Ith order approximations are concerned the amount of computational work is of the same magnitude as in the case $\beta = 0$ (both objects are at rest in the frame of observation). This is a consequence of the fact that time expression (40) contains two δ functions instead of one when $\beta = 0$, so that (54) [or (55)] involves not more than one integration (as is the case when $\beta = 0$).

V. THE SPECIAL CASE $\beta = 0$

Since the general formalism in the foregoing sections was developed by means of tensor equations it is valid in any frame of reference independently of whether both, only one, or none of the scatterers are at rest in the frame of observation.

This is in particular true for the integral equations (36) and (37) for the general far-field scattering amplitudes [or the separated equation (39) and its analog with

respect to the other scatterer]. [As to Eqs. (40) and (41) connecting the scattering amplitudes to the far-field components of the potentials and the electromagnetic field respectively they depend on the inertial frame as mentioned before.]

Let both scatterers be at rest in the frame of observation, i.e., $\beta = 0$. We can make use of (34), and from (36) we get

$$G_\lambda^{(1)}(k;\Psi) = G_{0\lambda}^{(1)}(k;\Psi) + (1/2\pi)^3 \int_{C_-} d^3\bar{p}\tilde{h}(\bar{p}, k_4)$$

$$\cdot \sum_{\kappa=1}^{2} g_{0\lambda}^{(1)}(\bar{k};\bar{p}, k_4, \kappa) G_\mu^{(2)}(\bar{p}, k_4;\Psi) b_\mu^{(\kappa)}(\bar{p}, k_4). \quad (56)$$

This equation and the analogous integral equation corresponding to (37) determine in principle the general far-field scattering amplitudes if the responses from the isolated obstacles corresponding to the exciting signal Ψ_λ as well as to an arbitrary plane wave are known. The field components can be calculated by means of (41).

These results may be compared with existing works. Equation (56) corresponds to Eq. (122) of Ref. 7, though the two equations are not identical which is due to several circumstances:

The analysis of Ref. 7 is based on the Helmholtz wave equation involving the time independent Green's function. In the present work our point of departure is the time dependent Green's function. In order to compare the results we have to assume that the exciting field Ψ_λ, which is quite arbitrary in (56), is a plane wave as in Ref. 7.

So, let Ψ_λ be given by (30). Equation (34) can be applied to (56). Since $k_4 = i|\bar{q}|$, where $\bar{q}$ is the wave vector and q_4 the normalized frequency of the incident plane wave, we supress notationally the dependence on the frequency. Furthermore, $g_4^{(\sigma)} = 0$ which follows from the Lorentz condition $\bar{g}^{(\sigma)} \cdot \bar{k} + g_4 k_4 = 0$ in connection with the relation $\bar{g}^{(\sigma)} \cdot \bar{k} = 0$ for the far-field. We obtain

$$\bar{g}^{(1)}(\bar{k};\bar{q},\bar{\epsilon}) = \bar{g}_0^{(1)}(\bar{k};\bar{q},\bar{\epsilon}) + \left(\frac{1}{2\pi}\right)^3 \int_{C_-} d^3\bar{p}\tilde{h}(p)\Big|_{p_4 = i|\bar{q}|}$$

$$\times \sum_{\kappa=1}^{2} \bar{g}_0^{(1)}(\bar{k};\bar{p},\kappa)[\bar{g}^{(2)}(\bar{p};\bar{q},\bar{\epsilon}) \cdot \bar{b}^{(\kappa)}(\bar{p})], \quad (57)$$

where $\bar{\epsilon}$ is the polarization vector of the incident plane wave.

Equation (41) reduces to

$$F_{\lambda\mu}^{(\sigma)}(\bar{x};\bar{q},\bar{\epsilon}) \simeq i2\pi cf(r)e^{i\bar{k}\cdot\bar{x}}[k_\mu g_\lambda^{(\sigma)}(\bar{k};\bar{q},\bar{\epsilon})$$

$$- k_\lambda g_\mu^{(\sigma)}(\bar{k};\bar{q},\bar{\epsilon})]_{k_4 = i|\bar{q}|} \quad (58)$$

Another difference between the present work and that of Ref. 7 is that we use the electromagnetic potentials. So, let us look at the electric field vector, for example. The components are given by F_{4k}/i, where $k \in \{1,2,3\}$, cf. Ref. 6, p. 217. Remembering that $g_4^{(\sigma)} = 0$ and that $f(r) = 1/4\pi|\bar{x}|$ for the three-dimensional case we obtain

$$\bar{E}^{(\sigma)}(\bar{x};\bar{q},\bar{\epsilon}) \simeq -\frac{i}{2}c|\bar{q}|\frac{e^{i\bar{k}\cdot\bar{x}}}{|\bar{x}|}\bar{g}^{(\sigma)}(\bar{k};\bar{q},\bar{\epsilon}). \quad (59)$$

In (58) and (59) $\bar{k} = |\bar{q}|\bar{x}/|\bar{x}|$, cf. (40).

Equation (59) agrees with the corresponding equation (116) of Ref. 7. In fact, if the electromagnetic potentials are introduced some manipulation shows that the surface integral [defined by (12), Ref. 7] in (116) reduces to an expression proportional to $c|\bar{q}|\bar{g}^{(\sigma)}$, where the defining surface integral for $\bar{g}^{(\sigma)}$ is given by (8).

Returning to Eq. (57) we can carry out the integration with respect to p_2 by means of the method of residues. Furthermore, we make use of (45), so that $\bar{g}^{(1)}$ and $\bar{g}^{(2)}$ refer to two systems of "local coordinates" with origins in the "centers" of the two scatterers respectively. (The two origins are connected by the vector $\bar{d}$). We obtain

$$\bar{g}^{(1)}(\bar{k};\bar{q},\bar{\epsilon}) = \bar{g}_0^{(1)}(\bar{k};\bar{q},\bar{\epsilon}) - \frac{i}{2}\left(\frac{1}{2\pi}\right)^2 \iint_{-\infty}^{\infty} dp_1 dp_3 \frac{e^{i(\bar{q}-\bar{p})\cdot\bar{d}}}{p_{2-}}$$

$$\times \sum_{\mu=1}^{2} \bar{g}_0^{(1)}(\bar{k};\bar{p},\kappa)[\bar{g}^{(2)}(\bar{p};\bar{q},\bar{\epsilon}) \cdot \bar{b}^{(\kappa)}(\bar{p})]. \quad (60)$$

This equation corresponds to Eq. (122), Ref. 7, and by means of some manipulation it is possible to show agreement if the following formal deviations are observed:

As to the summation signs they have a different meaning in the two equations in question. On one hand we can reformulate Eq. (60) in such a way that it concerns more than two objects in quite the same way as in (122), Ref. 7. On the other hand basic vectors may be introduced into (122), Ref. 7 in quite the same way as in Eq. (60). Then, two summations would appear in both equations in agreement with each other.

The two integration variables in (60) are real valued whereas both integration variables in (122), Ref. 7 are complex valued. But, it is not difficult to show by means of suitable substitutions that the equations are also in agreement with each other on this point.

[1] J. E. Burke and V. Twersky, Radio Sci. **68D**, 500—510 (1964).

[2] D. Censor, Radio Sci. **7**, 940—54 (1972).

[3] V. Twersky, *Scattering of Waves by Two Objects. Electromagnetic Waves*, edited by R. E. Langer (University of Wisconsin Press, Madison, Wisconsin, 1962).

[4] V. Twersky, J. Math. Phys. **3**, 83—91 (1962).

[5] Morse Feshbach, *Methods of Theoretical Physics, Part I*, (McGraw-Hill, New York, 1953), p. 838.

[6] A. Sommerfeld, *Electrodynamics* (Academic, New York, 1952), pp. 214—218.

[7] V. Twersky, J. Math. Phys. **8**, 589—610 (1967).

1428   J. Math. Phys., Vol. 17, No. 8, August 1976

S. Berntsen and G. Johannsen   1428

Nonspreading solutions to a class of differential equations*†

S. C. Chang‡ and Allen I. Janis

Department of Physics, University of Pittsburgh, Pittsburgh, Pennsylvania 15260
(Received 8 March 1976)

Explicit nonspreading (i.e., characteristically propagating) solutions to a certain class of two-dimensional, hyperbolic, differential equations are found. This class of equations is a generalization of equations that arose in a study of radiation in cosmological backgrounds.

1. INTRODUCTION

The propagation of disturbances of many physical systems can be described mathematically by second-order, normal-hyperbolic, differential equations. The solutions to these equations can be classified into two types: nonspreading, or characteristically propagating, solutions and spreading solutions.[1] A nonspreading solution describes a disturbance propagated at some characteristic speed, which is determined solely by the medium in which it propagates. In the case of a spreading solution, the front of the disturbance described by it will still propagate at the characteristic speed, but the wave will have a tail that travels at smaller speeds.

In the course of a study[2] of radiation in Friedmann-metric backgrounds, a class of two-dimensional, hyperbolic, differential equations was found to admit nonspreading solutions. (The equations describing radiation in Friedmann backgrounds were special cases of this wider class of equations.) In Sec. 2 we show how to construct these solutions. Some remarks concerning the associated substitution sequences, as defined by Kundt and Newman,[1] are given in Sec. 3.

2. NONSPREADING SOLUTIONS

First of all, we shall discuss the differential equations

$$\frac{\partial^2 \phi_{ll'}}{\partial v \partial v'} = \left[\frac{l(l+1)}{b^2} \operatorname{csch}^2\left(\frac{v+v'}{b}\right) - \frac{l'(l'+1)}{d^2} \operatorname{csch}^2\left(\frac{v-v'}{d}\right)\right]\phi_{ll'}. \tag{2.1}$$

Here $l, l' = 0, 1, 2, \cdots$; b and d are arbitrary constants, and $\phi_{ll'}$ are functions of v and v'. To aid in constructing solutions for Eqs. (2.1), we define

$$\tilde{\phi}_{ll'mm'} \equiv A_{ll'mm'}(v) \coth^m\left(\frac{v+v'}{b}\right) \coth^{m'}\left(\frac{v-v'}{d}\right), \tag{2.2}$$

with $m, m' = 0, 1, 2, \cdots$, and $A_{ll'mm'}(v)$ arbitrary functions of v. Then it is easy to see that

$$\frac{\partial^2 \tilde{\phi}_{ll'mm'}}{\partial v \partial v'} = A_{ll'mm'}(v) \coth^m\left(\frac{v+v'}{b}\right) \coth^{m'} \frac{v-v'}{d}$$

$$\times \left[\frac{m(m+1)}{b^2} \operatorname{csch}^2\left(\frac{v+v'}{b}\right) - \frac{m'(m'+1)}{d^2} \operatorname{csch}^2\left(\frac{v-v'}{d}\right)\right]$$

$$- \frac{m}{b}\dot{A}_{ll'mm'}(v) \coth^{m-1}\left(\frac{v+v'}{b}\right) \coth^{m'}\left(\frac{v-v'}{d}\right)$$

$$\times \operatorname{csch}^2\left(\frac{v+v'}{b}\right) + \frac{m'}{d}\dot{A}_{ll'mm'}(v) \coth^m\left(\frac{v+v'}{b}\right)$$

$$\times \coth^{m'-1}\left(\frac{v-v'}{d}\right) \operatorname{csch}^2\left(\frac{v-v'}{d}\right) - \frac{m(m-1)}{b^2}$$

$$\times A_{ll'mm'}(v) \coth^{m-2}\left(\frac{v+v'}{b}\right) \coth^{m'}\left(\frac{v-v'}{d}\right)$$

$$\times \operatorname{csch}^2\left(\frac{v+v'}{b}\right) + \frac{m'(m'-1)}{d^2} A_{ll'mm'}(v) \coth^m\left(\frac{v+v'}{b}\right)$$

$$\times \coth^{m'-2}\left(\frac{v-v'}{d}\right) \operatorname{csch}^2\left(\frac{v-v'}{d}\right), \tag{2.3}$$

with $\dot{A}_{ll'mm'}(v) \equiv dA_{ll'mm'}(v)/dv$. With the help of Eqs. (2.3), it is easy to see that

$$\tilde{\phi}_{ll'} \equiv \sum_{m=0}^{l} \sum_{m'=0}^{l'} \tilde{\phi}_{ll'mm'} \tag{2.4}$$

are solutions of Eqs. (2.1) if the following conditions are satisfied:

$$[l(l+1) - m(m+1)]A_{ll'mm'} + (m+2)(m+1)A_{ll'(m+2)m'}$$
$$+ b(m+1)\dot{A}_{ll'(m+1)m'} = 0, \tag{2.5}$$

$$[l'(l'+1) - m'(m'+1)]A_{ll'mm'} + (m'+2)(m'+1)A_{ll'm(m'+2)}$$
$$+ d(m'+1)\dot{A}_{ll'm(m'+1)} = 0, \tag{2.6}$$

with $A_{ll'mm'} = 0$ for any of the following cases:

(i) $m > l$, (ii) $m < 0$, (iii) $m' > l'$, (iv) $m' < 0$. (2.7)

In order to streamline the construction of solutions, we define a set of polynomials $F_{lm}(x)$, where $l = 0, 1, 2, \cdots$ and $m = 0, 1, 2, \ldots, l$, by means of the equations

$$[l(l+1) - m(m+1)]F_{lm}(x) + (m+2)(m+1)F_{l(m+2)}(x)$$
$$+ (m+1)xF_{l(m+1)}(x) = 0, \tag{2.8}$$

and

$$F_{ll}(x) \equiv 1, \quad F_{lm}(x) \equiv 0 \quad \text{for } m > l \text{ or } m < 0. \tag{2.9}$$

The self-consistency of conditions (2.8) and (2.9) may easily be verified. For $0 \le m < l$, Eqs. (2.8) define F_{lm} in terms of the F_{ll} defined in (2.9).

We can now see from the definition of the polynomials F_{lm} that

$$A_{ll'mm'}(v) \equiv F_{lm}(b\tilde{D})F_{l'm'}(d\tilde{D})A_{ll'}(v), \tag{2.10}$$

with $A_{ll'}(v)$ arbitrary functions of v and $\tilde{D} \equiv \partial/\partial v$, satisfy conditions (2.5), (2.6), and (2.7). Thus we conclude that

$$\tilde{\phi}_{ll'} \equiv \sum_{m=0}^{l} \sum_{m'=0}^{l'} [F_{lm}(b\tilde{D})F_{l'm'}(d\tilde{D})A_{ll'}(v)]$$

$$\times \coth^m\left(\frac{v+v'}{b}\right) \coth^{m'}\left(\frac{v-v'}{d}\right), \tag{2.11}$$

TABLE I. Incoming solutions, of the form $\tilde{\phi}_{ll'} = \sum\limits_{m=0}^{l}\sum\limits_{m'=0}^{l'}\tilde{\phi}_{ll'mm'}$, to Eqs. (2.13) for various choices of $R(v,v')$.

$R(v,v')$	$\tilde{\phi}_{ll'mm'}$
$\frac{l(l+1)}{b^2}\operatorname{csch}^2\left(\frac{v+v'}{b}\right) - \frac{l'(l'+1)}{d^2}\csc^2\left(\frac{v-v'}{d}\right)$	$\left[F_{lm}(b\tilde{D})G_{l'm'}(d\tilde{D})A_{ll'}(v)\right]\coth^m\left(\frac{v+v'}{b}\right)\cot^{m'}\left(\frac{v-v'}{d}\right)$
$\frac{l(l+1)}{b^2}\operatorname{csch}^2\left(\frac{v+v'}{b}\right) - \frac{l'(l'+1)}{d^2}\operatorname{csch}^2\left(\frac{v-v'}{d}\right)$	$\left[F_{lm}(b\tilde{D})F_{l'm'}(d\tilde{D})A_{ll'}(v)\right]\coth^m\left(\frac{v+v'}{b}\right)\coth^{m'}\left(\frac{v-v'}{d}\right)$
$\frac{l(l+1)}{b^2}\operatorname{csch}^2\left(\frac{v+v'}{b}\right) + \frac{l'(l'+1)}{d^2}\operatorname{sech}^2\left(\frac{v-v'}{d}\right)$	$\left[F_{lm}(b\tilde{D})F_{l'm'}(d\tilde{D})A_{ll'}(v)\right]\coth^m\left(\frac{v+v'}{b}\right)\tanh^{m'}\left(\frac{v-v'}{d}\right)$
$\frac{l(l+1)}{b^2}\operatorname{csch}^2\left(\frac{v+v'}{b}\right) - \frac{l'(l'+1)}{(v-v')^2}$	$\left[F_{lm}(b\tilde{D})H_{l'm'}(\tilde{D})A_{ll'}(v)\right]\coth^m\left(\frac{v+v'}{b}\right)[(v-v')^{-m'}]$
$\frac{l(l+1)}{b^2}\operatorname{csch}^2\left(\frac{v+v'}{b}\right) - \frac{l'(l'+1)}{d^2}\sec^2\left(\frac{v-v'}{d}\right)$	$\left[F_{lm}(b\tilde{D})G_{l'm'}(-d\tilde{D})A_{ll'}(v)\right]\coth^m\left(\frac{v+v'}{b}\right)\tan^{m'}\left(\frac{v-v'}{d}\right)$
$\frac{l(l+1)}{(v+v')^2} - \frac{l'(l'+1)}{(v-v')^2}$	$\left[H_{lm}(\tilde{D})H_{l'm'}(\tilde{D})A_{ll'}(v)\right](v+v')^{-m}(v-v')^{-m'}$
$\frac{l(l+1)}{(v+v')^2} + \frac{l'(l'+1)}{d^2}\operatorname{sech}^2\left(\frac{v-v'}{d}\right)$	$\left[H_{lm}(\tilde{D})F_{l'm'}(d\tilde{D})A_{ll'}(v)\right](v+v')^{-m}\tanh^{m'}\left(\frac{v-v'}{d}\right)$
$\frac{l(l+1)}{(v+v')^2} - \frac{l'(l'+1)}{d^2}\csc^2\left(\frac{v-v'}{d}\right)$	$\left[H_{lm}(\tilde{D})G_{l'm'}(d\tilde{D})A_{ll'}(v)\right](v+v')^{-m}\cot^{m'}\left(\frac{v-v'}{d}\right)$
$\frac{l(l+1)}{(v+v')^2} - \frac{l'(l'+1)}{d^2}\sec^2\left(\frac{v-v'}{d}\right)$	$\left[H_{lm}(\tilde{D})G_{l'm'}(-d\tilde{D})A_{ll'}(v)\right](v+v')^{-m}\tan^{m'}\left(\frac{v-v'}{d}\right)$
$\frac{l(l+1)}{b^2}\csc^2\left(\frac{v+v'}{b}\right) - \frac{l'(l'+1)}{d^2}\csc^2\left(\frac{v-v'}{d}\right)$	$\left[G_{lm}(b\tilde{D})G_{l'm'}(d\tilde{D})A_{ll'}(v)\right]\cot^m\left(\frac{v+v'}{b}\right)\cot^{m'}\left(\frac{v-v'}{d}\right)$
$\frac{l(l+1)}{b^2}\csc^2\left(\frac{v+v'}{b}\right) + \frac{l'(l'+1)}{d^2}\operatorname{sech}^2\left(\frac{v-v'}{d}\right)$	$\left[G_{lm}(b\tilde{D})F_{l'm'}(d\tilde{D})A_{ll'}(v)\right]\cot^m\left(\frac{v+v'}{b}\right)\tanh^{m'}\left(\frac{v-v'}{d}\right)$
$\frac{l(l+1)}{b^2}\csc^2\left(\frac{v+v'}{b}\right) - \frac{l'(l'+1)}{d^2}\sec^2\left(\frac{v-v'}{d}\right)$	$\left[G_{lm}(b\tilde{D})G_{l'm'}(-d\tilde{D})A_{ll'}(v)\right]\cot^m\left(\frac{v+v'}{b}\right)\tan^{m'}\left(\frac{v-v'}{d}\right)$
$\frac{l(l+1)}{b^2}\sec^2\left(\frac{v+v'}{b}\right) - \frac{l'(l'+1)}{d^2}\sec^2\left(\frac{v-v'}{d}\right)$	$\left[G_{lm}(-b\tilde{D})G_{l'm'}(-d\tilde{D})A_{ll'}(v)\right]\tan^m\left(\frac{v+v'}{b}\right)\tan^{m'}\left(\frac{v-v'}{d}\right)$
$\frac{l(l+1)}{b^2}\sec^2\left(\frac{v+v'}{b}\right) + \frac{l'(l'+1)}{d^2}\operatorname{sech}^2\left(\frac{v-v'}{d}\right)$	$\left[G_{lm}(-b\tilde{D})F_{l'm'}(d\tilde{D})A_{ll'}(v)\right]\tan^m\left(\frac{v+v'}{b}\right)\tanh^{m'}\left(\frac{v-v'}{d}\right)$
$-\frac{l(l+1)}{b^2}\operatorname{sech}^2\left(\frac{v+v'}{b}\right) + \frac{l'(l'+1)}{d^2}\operatorname{sech}^2\left(\frac{v-v'}{d}\right)$	$\left[F_{lm}(b\tilde{D})F_{l'm'}(d\tilde{D})A_{ll'}(v)\right]\tanh^m\left(\frac{v+v'}{b}\right)\tanh^{m'}\left(\frac{v-v'}{d}\right)$

with $A_{ll'}(v)$ arbitrary functions of v, are solutions of Eqs. (2.1).

The functions $\tilde{\phi}_{ll'}$, as defined by Eqs. (2.11), are nonspreading solutions of Eqs. (2.1). This can be seen from the facts that v and v' label the characteristic surfaces of Eqs. (2.1) and that $\tilde{\phi}_{ll'} = 0$ if v is outside the support of the "news functions" $A_{ll'}(v)$.

Because Eqs. (2.1) are invariant under the interchange of v and v', it is easy to see that

$$\hat{\phi}_{ll'} \equiv \sum_{m=0}^{l}\sum_{m'=0}^{l'}[F_{lm}(b\hat{D})F_{l'm'}(d\hat{D})B_{ll'}(v')]$$

$$\times\coth^m\left(\frac{v+v'}{b}\right)\coth^{m'}\left(\frac{v'-v}{d}\right), \qquad (2.12)$$

with $B_{ll'}(v')$ arbitrary functions of v' and $\hat{D} \equiv \partial/\partial v'$, are also nonspreading solutions of Eqs. (2.1).

Solutions that contain arbitrary functions of v will be designated as "incoming solutions," while those containing arbitrary functions of v' will be designated as

"outgoing solutions." With this convention, $\tilde{\phi}_{ll'}$ and $\hat{\phi}_{ll'}$ are incoming and outgoing solutions, respectively. For equations that are invariant under the interchange of v and v', incoming and outgoing solutions can be obtained from each other by the interchange of v and v'.

Based on similar arguments, we can construct incoming and outgoing nonspreading solutions for a larger set of differential equations. These equations are all of the form

$$\frac{\partial^2 \phi_{ll'}}{\partial v\partial v'} = R(v,v')\phi_{ll'}, \qquad (2.13)$$

where $R(v,v')$ is a given function that is invariant under the interchange of v and v'. The incoming solutions, $\tilde{\phi}_{ll'}$, of these equations for a number of different choices of $R(v,v')$ are given in Table I. These solutions are expressed in terms of the three sets of polynomials $F_{lm}(x)$, $G_{lm}(x)$, and $H_{lm}(x)$, with $l = 0,1,2,\cdots$ and $m = 0,1,2,\ldots,l$, where the new polynomials $G_{lm}(x)$ and $H_{lm}(x)$ are defined by the equations

$$[l(l+1) - m(m+1)]G_{lm}(x) - (m+2)(m+1)G_{l(m+2)}(x)$$

$$+ (m+1)xG_{l(m+1)}(x) = 0, \tag{2.14}$$

$$G_{ll}(x) \equiv 1, \quad G_{lm}(x) \equiv 0 \quad \text{for} \quad m > l \quad \text{or} \quad m < 0, \tag{2.15}$$

$$[l(l+1) - m(m+1)]H_{lm}(x) + (m+1)xH_{l(m+1)}(x) = 0,$$

$$\tag{2.16}$$

$$H_{ll}(x) \equiv 1, \quad H_{lm}(x) \equiv 0 \quad \text{for} \quad m > l \quad \text{or} \quad m < 0. \tag{2.17}$$

Because Eqs. (2.13) are invariant under the interchange of v and v', the outgoing solutions, $\hat{\phi}_{ll'}$, can easily be obtained from the incoming solutions, $\tilde{\phi}_{ll'}$.

By means of the transformation $v' \to -v'$, we can construct solutions to certain partial differential equations that are not listed in Table I. For example, under the transformation $v' \to -v'$, the equations

$$\frac{\partial^2 \phi_{ll'}}{\partial v \partial v'} = \left[\frac{l(l+1)}{b^2} \csc^2\!\left(\frac{v+v'}{b}\right) \right.$$
$$\left. - \frac{l'(l'+1)}{d^2} \operatorname{csch}^2\!\left(\frac{v-v'}{d}\right) \right] \phi_{ll'} \tag{2.18}$$

become

$$\frac{\partial^2 \phi_{ll'}}{\partial v \partial v'} = \left[\frac{l'(l'+1)}{d^2} \operatorname{csch}^2\!\left(\frac{v+v'}{d}\right) \right.$$
$$\left. - \frac{l(l+1)}{b^2} \csc^2\!\left(\frac{v-v'}{b}\right) \right] \phi_{ll'}. \tag{2.19}$$

From the incoming solutions of Eqs. (2.19) given in Table I, it is easy to see that the incoming solutions to Eqs. (2.18) are

$$\tilde{\phi}_{ll'} = \sum_{m=0}^{l} \sum_{m'=0}^{l'} \left[F_{l'm'}(d\tilde{D})G_{lm}(b\tilde{D})A_{ll'}(v) \right]$$

$$\times \cot^m\!\left(\frac{v+v'}{b}\right) \coth^{m'}\!\left(\frac{v-v'}{d}\right). \tag{2.20}$$

The outgoing solutions of Eqs. (2.18) can be obtained from the incoming solutions by the interchange of v with v' and the replacement of $\tilde{D}$ by $\hat{D}$.

3. DISCUSSION

Kundt and Newman[1] have discussed the relation between the existence of nonspreading solutions to hyperbolic differential equations in two dimensions and the termination of what they have referred to as "substitution sequences." The solutions that we have presented in this paper are consistent with these results of Kundt and Newman.

As a result of the fact that $F_{lm}(x)$, $G_{lm}(x)$, and $H_{lm}(x)$ are polynomial functions of x, all of the incoming solutions discussed above are of the form

$$\tilde{\phi} = \sum_{k=0}^{n} g_k(v, v') \frac{d^{n-k}A(v)}{dv^{n-k}}, \tag{3.1}$$

with n some positive integer and $g_0 \neq 0$. It then follows[1,3] that the associated substitution sequences terminate to the right. Because the substitution sequences are symmetric, they also terminate to the left.

The substitution sequences associated with the differential equations discussed in this paper are not only symmetric and terminating, they are also nontrivial; i.e., the members of these sequences are not functions of $(v+v')$ or $(v-v')$ alone. The existence of these nontrivial sequences can be used to test the validity of certain conjectures concerning the analytic shape of members of terminating symmetric sequences.[4]

[1] W. Kundt and E.T. Newman, J. Math. Phys. **9**, 2193 (1968).
[2] S.C. Chang and A.I. Janis, J. Math. Phys. **17**, 1432 (1976).
[3] S.C. Chang, *Radiation in Cosmological Backgrounds*, Ph.D. dissertation (University of Pittsburgh, 1974).
[4] For example, Eqs. (11.22) of Ref. 1 are conjectured to be true for general terminating, symmetric, substitution sequences.

Radiation in cosmological backgrounds*†

S. C. Chang‡ and Allen I. Janis

Department of Physics, University of Pittsburgh, Pittsburgh, Pennsylvania 15260
(Received 8 March 1976)

The purpose of this investigation is to find the conditions for characteristic propagation of multipole radiation in Friedmann backgrounds. The radiation fields studied are Klein–Gordon scalar fields, conformally invariant scalar fields, electromagnetic fields, and gravitational fields. The behavior of electromagnetic and conformally invariant scalar radiation is similar to that of the corresponding radiation in flat space-time, since both fields satisfy conformally invariant equations and the Friedmann backgrounds are conformally flat. Thus characteristically propagating solutions are possible for both fields in any Friedmann background. For the Klein–Gordon and gravitational fields, it is found that characteristic propagation is possible only for special Friedmann backgrounds. Two physically important Friedmann backgrounds, those for which $P = 0$ and $P = \rho/3$ (where P is pressure and ρ is density), are among these special backgrounds for both types of radiation. In the course of this study, all Friedmann backgrounds for which $P = \alpha\rho$, where α is an arbitrary constant, are found; the methods used and the resulting solutions are much simpler than those previously given by Tauber.

1. INTRODUCTION

The propagation of waves in most media is not of the nonspreading, or characteristically propagating, type; i. e. , the waves do not propagate at some characteristic speed that is solely determined by the medium in which they propagate. Usually, the wavefront of a disturbance will propagate at some characteristic speed, but it will be followed by a wave tail that propagates at smaller speeds. This tail, in some sense, can be considered as arising from the continuous backscattering of the wave.

When radiation propagates in various space–time backgrounds, there is usually a radiation tail resulting from the interaction with the space–time curvature. Kundt and Newman[1] investigated the problem of backscattering for scalar and electromagnetic waves in both flat and Schwarzschild backgrounds. Their results indicated that, although the flat background produced no backscattering, there was always backscattering in the Schwarzschild background. The questions naturally arise then, of whether backscattering is unavoidable in curved backgrounds, and if not, under what circumstances radiation tails will be absent. In order to give a partial answer to these questions, we study radiation in Friedmann backgrounds (with vanishing cosmological constant). These backgrounds, which include models with negative, positive, and zero spatial curvature, are highly symmetrical and thus relatively easy to handle. A brief introduction to the Friedmann models is given in Appendix A. The radiation fields that we study are scalar, electromagnetic, and gravitational radiation. The basic equations governing their propagation in Friedmann backgrounds are discussed in Secs. 2, 3, and 4, respectively. The circumstances under which these equations admit nonspreading solutions are then discussed in Sec. 5.

Both conformally invariant scalar radiation[2] and electromagnetic radiation satisfy conformally invariant field equations. Since the Friedmann backgrounds are conformally flat, it is not difficult to see that the behavior of such radiation in Friedmann backgrounds is similar to that of the corresponding radiation in flat space–time, and can thus be described by nonspread-

ing solutions.[2-4] In.other words, radiation without backscattering is possible for any Friedmann background. The results of our study, as expected, confirm these conclusions.

The Klein–Gordon equation is a well known example of an equation that is not conformally invariant. We thus might expect that scalar radiation governed by this equation would usually be accompanied by backscattering when it propagates in Friedmann backgrounds. Again, our study confirms this expectation. We find, however, that multipole radiation of the Klein–Gordon field can have the property of characteristic propagation for a special class of Friedmann backgrounds. Among this special class are two physically important backgrounds, those with the equations of state $P = 0$ and $P = \rho/3$, where P and ρ are pressure and density, respectively.

The question of the existence of backscattering for gravitational radiation in Friedmann backgrounds is closely bound to the question of what field equations to choose as governing this radiation. Penrose[2] has suggested a set of conformally invariant equations for zero-rest-mass fields of spin s (where $s = \frac{1}{2}, 1, \frac{3}{2}, \cdots$), viz. ,

$$\nabla^{A\dot{Y}}\phi_{AB\cdots K} = 0, \tag{1.1}$$

where $\phi_{AB\cdots K}$ is a totally symmetric spinor with $2s$ indices. For the special case of $s = 2$, Eqs. (1.1) reduce to

$$\nabla^{A\dot{Y}}\phi_{ABCD} = 0. \tag{1.2}$$

One might think that Eqs. (1.2), with the attractive property of conformal invariance, would be the best choice of equations governing gravitational radiation in Friedmann backgrounds. On the other hand, the following argument seems to work against this choice. We first note that in a vacuum the Bianchi identities reduce to the form

$$\nabla^{A\dot{Y}}\psi_{ABCD} = 0, \tag{1.3}$$

where ψ_{ABCD} is the Weyl spinor. Since Eqs. (1.3) are the same as Eqs. (1.2), the choice of Eqs. (1.2) as the

governing equations for gravitational radiation propagating in a vacuum seems reasonable. However, in more general cases the Bianchi identities need not take the form (1.3). Thus, in the case of the Friedmann backgrounds, which are not vacuum backgrounds, the choice of Eqs. (1.2) as the governing equations is no longer very convincing.

The approach we take to discover the appropriate governing equations is through perturbation theory applied to the Einstein field equations, with the Friedmann solutions as the zero-order solutions. We employ the Newman—Penrose formalism,[5] which we consider to be the one best suited for our study. As we are only interested in gravitational radiation, we assume that the Ricci tensor has vanishing first-order perturbations (which could be considered as perturbations of the material, as opposed to the gravitational, background), while allowing nonvanishing perturbations of the Weyl tensor. Since the Weyl tensor now satisfies equations that are not conformally invariant, we again find that the multipole radiation can propagate characteristically only for a special class of Friedmann backgrounds. As in the Klein—Gordon case, we find, rather surprisingly, that the backgrounds with equations of state $P = 0$ and $P = \rho/3$ belong to this special class.

The results of Secs. 2—4 are summarized in Table I. It is seen there that all of the equations governing the various fields share the important property that their characteristic hypersurfaces are null hypersurfaces of the background space—time. In other words, the characteristic speed of all of the different types of radiation is the speed of light. It is also seen that the equations governing the Klein—Gordon and the gravitational radiation fields are related in a strange way. As explained in Sec. 5, they are the images of each other under a simple mapping of one Friedmann background into another of the same spatial-curvature type. Other surprising results, e.g., the relation between Tables II and III, which describe certain Friedmann backgrounds that lead to characteristically propagating multipole fields, are also discussed in Sec. 5, and summarized in Sec. 6.

Finally, we note that Friedmann backgrounds satisfying the equation of state $P = \alpha \rho$, where α is an arbitrary constant, are discussed fully in Appendix B. Our method of finding all such backgrounds, and their resulting forms, are much simpler than those given previously by Tauber.[6]

2. SCALAR RADIATION

Our discussion of radiation in Friedmann backgrounds will start with the simplest type, i.e., scalar radiation. Specifically, we shall discuss those scalar fields that satisfy an equation of the form

$$(g^{\mu\nu}\psi_{,\nu})_{;\mu} + \frac{(1-n)}{6} R\psi = 0, \qquad (2.1)$$

where n is an arbitrary constant. We note that for $n = 1$ and $n = 0$, Eq. (2.1) reduces to the Klein—Gordon equation and the conformally invariant scalar field equation,[2] respectively.

If we take ψ to be a first-order quantity and the metric backgrounds to be those of Eqs. (A8)—(A10), a straightforward calculation shows that the linearized version of Eq. (2.1) can be transformed into

$$\frac{\partial^2 \phi}{\partial v \, \partial v'} = -\left(na^2 \underset{0}{\Lambda} + \frac{I}{4}\,\delta\bar{\delta}\right)\phi, \qquad (2.2)$$

where ϕ is defined by

$$\phi \equiv e^H \psi, \qquad (2.3)$$

and H, $\underset{0}{\Lambda}$, and I are defined by Eqs. (A11), (A21), and (A23), respectively. The differential operators δ and $\bar{\delta}$ are essentially the spin-weighted differentiation operators on a sphere.[7,8] If we expand ϕ as

$$\phi = \sum_{l=0}^{\infty} \sum_{m=-l}^{l} \phi_{lm}(v,v') Y_{lm}(\zeta,\bar{\zeta}), \qquad (2.4)$$

where $Y_{lm}(\zeta,\bar{\zeta})$ are the spherical harmonics, then Eq. (2.2) implies that

$$\frac{\partial^2 \phi_{lm}}{\partial v \, \partial v'} = \left[\frac{l(l+1)I}{4} - na^2 \underset{0}{\Lambda}\right]\phi_{lm} \qquad (2.5)$$

for $l = 0, 1, 2, \cdots$ and $m = -l, -l+1, \ldots, l$.

Equations (2.5) describe the multipole radiation, in Friedmann backgrounds, for scalar fields governed by Eq. (2.1). In particular, for conformally invariant and Klein—Gordon scalar fields, Eqs. (2.5) become

$$\frac{\partial^2 \phi_{lm}}{\partial v \, \partial v'} = \frac{l(l+1)I}{4}\,\phi_{lm} \qquad (2.6)$$

and

$$\frac{\partial^2 \phi_{lm}}{\partial v \, \partial v'} = \left(\frac{l(l+1)I}{4} - a^2 \underset{0}{\Lambda}\right)\phi_{lm}, \qquad (2.7)$$

respectively. Equations (2.6) and (2.7) will serve as the basic equations for our study of characteristic scalar radiation in Sec. 5.

3. ELECTROMAGNETIC RADIATION

We shall employ the Newman—Penrose formalism[5] to study electromagnetic radiation in Friedmann backgrounds. Maxwell's equations in this formalism are[5]

$$D\Phi_1 - \bar{\delta}\Phi_0 = (\pi - 2\alpha)\Phi_0 + 2\rho\Phi_1 - \kappa\Phi_2, \qquad (3.1)$$

$$D\Phi_2 - \bar{\delta}\Phi_1 = -\lambda\Phi_0 + 2\pi\Phi_1 + (\rho - 2\epsilon)\Phi_2, \qquad (3.2)$$

$$\delta\Phi_1 - \Delta\Phi_0 = (\mu - 2\gamma)\Phi_0 + 2\tau\Phi_1 - \sigma\Phi_2, \qquad (3.3)$$

$$\delta\Phi_2 - \Delta\Phi_1 = -\nu\Phi_0 + 2\mu\Phi_1 + (\tau - 2\beta)\Phi_2. \qquad (3.4)$$

In linearizing these equations, we take the field variables Φ_0, Φ_1, and Φ_2 to be first-order quantities. Then only the zero-order terms of the other quantities need to be considered, and the resulting linearized equations describe electromagnetic fields in the given background. If we adopt Hawking's tetrad and coordinate conditions,[9] and use the background quantities given in Appendix A, it is easy to obtain and solve the linearized versions of Eqs. (3.1)—(3.4).[7] The entire solution can be expressed in terms of certain initial data and the spin-weight-zero quantity Φ_1. We find that Φ_1 obeys the equation

$$\frac{\partial^2 \phi}{\partial v \, \partial v'} = -\frac{I}{4}\,\delta\bar{\delta}\phi, \qquad (3.5)$$

where ϕ is defined by

$$\phi \equiv e^{2H}\Phi_1. \tag{3.6}$$

If we consider the expansion of ϕ in spherical harmonics,

$$\phi = \sum_{l=0}^{\infty} \sum_{m=-l}^{l} \phi_{lm}(v, v') Y_{lm}(\zeta, \bar{\zeta}), \tag{3.7}$$

then Eq. (3.5) implies that

$$\frac{\partial^2 \phi_{lm}}{\partial v \, \partial v'} = \frac{l(l+1)I}{4} \phi_{lm} \tag{3.8}$$

for $l = 0, 1, 2, \cdots$ and $m = -l, -l+1, \ldots, l$.

Equations (3.8) are the basic equations describing electromagnetic multipole radiation in Friedmann backgrounds. These equations will be studied further in Sec. 5.

4. GRAVITATIONAL RADIATION

Again, the formalism of Newman and Penrose[5] will be employed to attack the problem of gravitational radiation in Friedmann backgrounds. We specialize this formalism by using Hawking's notation, definitions, tetrad and coordinate conditions, and field equations,[9] which will not be reproduced here.

All quantities are separated into unperturbed and perturbed parts, e.g., Φ_{00} is written as $\Phi_{00} = \underset{0}{\Phi_{00}} + \underset{1}{\Phi_{00}}$, with $\underset{0}{\Phi_{00}}$ the unperturbed part and $\underset{1}{\Phi_{00}}$ the perturbed part. We consider only pure gravitational radiation with the perturbed parts of all components of the Ricci tensor vanishing. Thus we have

$$\underset{1}{\Phi_{00}} = \underset{1}{\Phi_{11}} = \underset{1}{\Phi_{22}} = \underset{1}{\Phi_{01}} = \underset{1}{\Phi_{02}} = \underset{1}{\Phi_{12}} = \underset{1}{\Lambda} = 0. \tag{4.1}$$

Conditions (4.1), together with the unperturbed (i.e., zero-order) quantities given in Appendix A, can be used to obtain the linearized forms of the field equations, and the resulting equations can be solved.[7] The entire solution can be expressed in terms of certain initial data and the spin-weight-zero spin component of the Weyl tensor, Ψ_2 (which is equal to $\underset{1}{\Psi_2}$, because $\underset{0}{\Psi_2} = 0$). We find that Ψ_2 obeys the equation

$$\frac{\partial^2 \phi}{\partial v \, \partial v'} = -\left(\underset{0}{\Phi_{22}} + \frac{I}{4}\delta\bar{\delta}\right)\phi, \tag{4.2}$$

where ϕ is defined by

$$\phi \equiv e^{3H}\Psi_2. \tag{4.3}$$

When $\underset{0}{\Phi_{22}} \neq 0$, Ψ_2 is further restricted by the condition

$$\Psi_2 + \bar{\Psi}_2 = 0; \tag{4.4}$$

in other words, Ψ_2 has only a magnetic part.[8] If we consider the expansion of ϕ in spherical harmonics,

$$\phi = \sum_{l=0}^{\infty} \sum_{m=-l}^{l} \phi_{lm}(v, v') Y_{lm}(\zeta, \bar{\zeta}), \tag{4.5}$$

then Eqs. (4.2) and (4.4) imply that

$$\frac{\partial^2 \phi_{lm}}{\partial v \, \partial v'} = \left(\frac{l(l+1)I}{4} - \underset{0}{\Phi_{22}}\right)\phi_{lm} \tag{4.6}$$

and

$$\phi_{lm} + (-1)^m \phi_{l(-m)} = 0, \tag{4.7}$$

respectively, for $l = 0, 1, 2, \cdots$ and $m = -l, -l+1, \ldots, l$.

These equations, which also will be studied further in Sec. 5, are our basic equations for gravitational multipole radiation in Friedmann backgrounds.

It may be seen from Eqs. (4.6) that v and v' label the characteristic hypersurfaces of these equations. As noted in Appendix A, v and v' also label null hypersurfaces in the Friedmann backgrounds. Thus we see that the wavefronts of the multipole radiation governed by Eqs. (4.6) propagate with the speed of light. It is easily seen that the analogous conclusions hold for the scalar and electromagnetic radiation fields discussed in Secs. 2 and 3.

5. CHARACTERISTIC RADIATION

In Secs. 2–4, we found that the study of scalar, electromagnetic, and gravitational radiation in Friedmann backgrounds could be reduced to the study of certain two-dimensional differential equations. In the present section, we shall discuss the circumstances under which these equations admit characteristically propagating (i.e., nonspreading) radiative solutions.[10]

First, a summary of the major results of the preceding sections will be useful. This summary is given in Table I, where Eqs. (5.1)—(5.4) are the equations

TABLE I. Differential equations governing radiation in Friedmann backgrounds.

Type of field	Field Variable	Definition of ϕ	Equation for ϕ		Equation for ϕ_{lm}	
Conformally Invariant Scalar Field	ψ	$\phi \equiv e^{H}\psi$	$\dfrac{\partial^2 \phi}{\partial v \partial v'} = -\dfrac{I}{4}\delta\bar{\delta}\phi$	(5.1)	$\dfrac{\partial^2 \phi_{lm}}{\partial v \partial v'} = \dfrac{l(l+1)I}{4}\phi_{lm}$	(5.5)
Klein–Gordon Scalar Field	ψ	$\phi \equiv e^{H}\psi$	$\dfrac{\partial^2 \phi}{\partial v \partial v'} = -\left(a^2\underset{0}{\Lambda} + \dfrac{I}{4}\delta\bar{\delta}\right)\phi$	(5.2)	$\dfrac{\partial^2 \phi_{lm}}{\partial v \partial v'} = \left(\dfrac{l(l+1)I}{4} - a^2\underset{0}{\Lambda}\right)\phi_{lm}$	(5.6)
Electromagnetic Field	Φ_1	$\phi \equiv e^{2H}\Phi_1$	$\dfrac{\partial^2 \phi}{\partial v \partial v'} = -\dfrac{I}{4}\delta\bar{\delta}\phi$	(5.3)	$\dfrac{\partial^2 \phi_{lm}}{\partial v \partial v'} = \dfrac{l(l+1)I}{4}\phi_{lm}$	(5.7)
Gravitational Field	Ψ_2	$\phi \equiv e^{3H}\Psi_2$	$\dfrac{\partial^2 \phi}{\partial v \partial v'} = -\left(\underset{0}{\Phi_{22}} + \dfrac{I}{4}\delta\bar{\delta}\right)\phi$	(5.4)	$\dfrac{\partial^2 \phi_{lm}}{\partial v \partial v'} = \left(\dfrac{l(l+1)I}{4} - \underset{0}{\Phi_{22}}\right)\phi_{lm}$	(5.8)

determining the basic field variables and Eqs. (5.5)–(5.8) are the corresponding two-dimensional equations obtained by an expansion in spherical harmonics. We see immediately from this table that Eqs. (5.1) and (5.3) are identical, as are, correspondingly, Eqs. (5.5) and (5.7). Also, these equations are independent of the specific form of the function $a(t)$, as is evident from Eq. (A23). These properties, of course, are not accidental. They result from the facts that these equations are associated with conformally invariant fields and that the Friedmann backgrounds are conformally flat. For the Klein–Gordon and gravitational fields, the equations are more complex and depend on the specific form of $a(t)$. Although Eqs. (5.2) and (5.4) are not identical, they are still very similar, and in fact have a strange connection between them. In order to see this, we first define

$$s(a) \equiv a^2 \Lambda_0. \tag{5.9}$$

We then have, from Eqs. (A21) and (A22), that

$$s(a) = \begin{cases} a''/4a - \frac{1}{4}, \\ a''/4a + \frac{1}{4}, \\ a''/4a, \end{cases} \tag{5.10}$$

and

$$\Phi_{22}(a) = \begin{cases} (a')^2/2a^2 - a''/4a - \frac{1}{4}, \\ (a')^2/2a^2 - a''/4a + \frac{1}{4}, \\ (a')^2/2a^2 - a''/4a, \end{cases} \tag{5.11}$$

for models with negative, positive, and zero spatial curvature, respectively. It follows from Eqs. (5.10) and (5.11) that

$$s(1/a) = \Phi_{22}(a). \tag{5.12}$$

Thus $s(a)$ and $\Phi_{22}(a)$ are the images of each other under the transformation $a \to 1/a$ of backgrounds. It is now easy to see that Eqs. (5.2) and (5.4) are the images of each other under the same background transformation, and so are Eqs. (5.6) and (5.8).

We now embark on the study of Eqs. (5.5)–(5.8). Since the functions ϕ_{lm} satisfy the same equations for every allowed value of m, we shall drop the index m in the following treatment.

A. Equations (5.5) and (5.7)

With the aid of Eqs. (A23), it is seen that Eqs. (5.5) and (5.7) can be rewritten as

$$\frac{\partial^2 \phi_l}{\partial v \, \partial v'} = \frac{l(l+1)}{4} \operatorname{csch}^2\left(\frac{v+v'}{2}\right)\phi_l, \tag{5.13}$$

$$\frac{\partial^2 \phi_l}{\partial v \, \partial v'} = \frac{l(l+1)}{4} \csc^2\left(\frac{v+v'}{2}\right)\phi_l, \tag{5.14}$$

$$\frac{\partial^2 \phi_l}{\partial v \, \partial v'} = \frac{l(l+1)}{(v+v')^2}\phi_l, \tag{5.15}$$

for models with negative, positive, and zero spatial curvature, respectively. Equations (5.13)–(5.15) are special cases, obtained by setting $b=2$ and $l'=0$, of the equations

$$\frac{\partial^2 \phi_{ll'}}{\partial v \, \partial v'} = \left[\frac{l(l+1)}{b^2} \operatorname{csch}^2\left(\frac{v+v'}{b}\right) - \frac{l'(l'+1)}{d^2} \operatorname{csch}^2\left(\frac{v-v'}{d}\right)\right] \times \phi_{ll'}, \tag{5.16}$$

$$\frac{\partial^2 \phi_{ll'}}{\partial v \, \partial v'} = \left[\frac{l(l+1)}{b^2} \csc^2\left(\frac{v+v'}{b}\right) - \frac{l'(l'+1)}{d^2} \csc^2\left(\frac{v-v'}{d}\right)\right]\phi_{ll'}, \tag{5.17}$$

$$\frac{\partial^2 \phi_{ll'}}{\partial v \, \partial v'} = \left[\frac{l(l+1)}{(v+v')^2} - \frac{l'(l'+1)}{(v-v')^2}\right]\phi_{ll'}, \tag{5.18}$$

respectively. In the preceding paper,[10] it was shown that Eqs. (5.16)–(5.18) admit nonspreading solutions for all nonnegative integral values of l and l' and arbitrary constant values of b and d. Thus we conclude that Eqs. (5.5) and (5.7) always admit nonspreading solutions.

B. Equations (5.6)

We initiate the study of this case by considering those Friedmann backgrounds that satisfy the equation of state

$$P = \alpha \rho, \tag{5.19}$$

where α is an arbitrary constant, and P and ρ are pressure and density, respectively.

In those cases for which $3\alpha + 1 \neq 0$, it is shown in Appendix B that $a^2 \Lambda_0$ for this class of backgrounds is given by

$$a^2 \Lambda_0 = \begin{cases} \dfrac{1-3\alpha}{8} \operatorname{csch}^2\left(\dfrac{3\alpha+1}{2}t\right), \\[2ex] \dfrac{1-3\alpha}{8} \csc^2\left(\dfrac{3\alpha+1}{2}t\right), \\[2ex] \dfrac{1-3\alpha}{2(3\alpha+1)^2}\dfrac{1}{t^2}, \end{cases} \tag{5.20}$$

for models with negative, positive, and zero spatial curvature, respectively. Using Eqs. (A23), (5.20), and (A7), we can rewrite Eqs. (5.6) as

$$\frac{\partial^2 \phi_l}{\partial v \, \partial v'} = \left[\frac{l(l+1)}{4} \operatorname{csch}^2\left(\frac{v+v'}{2}\right) - \frac{1-3\alpha}{8} \times \operatorname{csch}^2\left(\frac{3\alpha+1}{4}(v-v')\right)\right]\phi_l, \tag{5.21}$$

$$\frac{\partial^2 \phi_l}{\partial v \, \partial v'} = \left[\frac{l(l+1)}{4} \csc^2\left(\frac{v+v'}{2}\right) - \frac{1-3\alpha}{8} \csc^2\left(\frac{3\alpha+1}{4}(v-v')\right)\right] \times \phi_l, \tag{5.22}$$

$$\frac{\partial^2 \phi_l}{\partial v \, \partial v'} = \left(\frac{l(l+1)}{(v+v')^2} - \frac{2(1-3\alpha)}{(3\alpha+1)^2}\frac{1}{(v-v')^2}\right)\phi_l, \tag{5.23}$$

for models with negative, positive, and zero spatial curvature, respectively.

A sufficient condition that Eqs. (5.21)–(5.23) admit nonspreading solutions is that these equations be special cases of Eqs. (5.16)–(5.18), respectively, where, in the latter equations, b and d are constants and l and l' are nonnegative integers. This will be so provided that d is chosen to be

TABLE II. Values of α satisfying Eqs. (5.26) for allowed values of l'.

$l'=0$	$l'=1$	$l'=2$	$l'=3$	---	$l'\to\infty$
$\alpha=\frac{1}{3}$	$\alpha=0$	$\alpha=-\frac{1}{9}$	$\alpha=-\frac{1}{6}$	---	
					$\alpha\to-\frac{1}{3}$
	$\alpha=-1$	$\alpha=-\frac{2}{3}$	$\alpha=-\frac{5}{9}$	---	

$$d=\pm 4/(3\alpha+1),\qquad (5.24)$$

and that, for allowed values of l', α and l' are related by

$$(3\alpha+1)^2 l'(l'+1)=2(1-3\alpha).\qquad (5.25)$$

The solutions of Eqs. (5.25) for α are

$$\alpha=\frac{1-l'}{3(1+l')},\quad l'=0,1,2,\cdots,\qquad (5.26a)$$

and

$$\alpha=-\frac{l'+2}{3l'},\quad l'=1,2,\cdots.\qquad (5.26b)$$

The numerical results of Eqs. (5.26) are indicated in Table II.

Among the Friedmann backgrounds specified by Eqs. (5.26) (or Table II), only those with $\alpha=0$ and $\alpha=\frac{1}{3}$ have nonnegative pressure. For $\alpha=0$, the background is pressure free. For $\alpha=\frac{1}{3}$, the pressure is entirely due to radiation.

Let us now drop the requirement that the Friedmann backgrounds be subject to the equation of state (5.19). We then see, making use of Eqs. (5.9), (5.10), and (A7), that Eqs. (5.6) will still be special cases of Eqs. (5.16)—(5.18) provided that

$$\frac{a''}{4a}-\frac{1}{4}=\frac{l'(l'+1)}{d^2}\,\text{csch}^2\!\left(\frac{2t}{d}\right),\qquad (5.27)$$

$$\frac{a''}{4a}+\frac{1}{4}=\frac{l'(l'+1)}{d^2}\,\csc^2\!\left(\frac{2t}{d}\right),\qquad (5.28)$$

$$\frac{a''}{4a}=\frac{l'(l'+1)}{4t^2},\qquad (5.29)$$

for models with negative, positive, and zero spatial curvature, respectively. Equations (5.27)—(5.29) are not difficult to solve. For example, if we define $x\equiv\coth(2t/d)$, then Eq. (5.27) becomes

$$(1-x^2)\frac{d^2a}{dx^2}-2x\frac{da}{dx}+\left(l'(l'+1)-\frac{(d/2)^2}{1-x^2}\right)a=0,\quad (5.27')$$

which is one form of Legendre's equation and is thus easily solved.

We have found that, for a variety of different Friedmann backgrounds, including some physically important cases, the equations for multipole radiation of the Klein—Gordon field become special cases of Eqs. (5.16)—(5.18) and thus admit nonspreading solutions. In the preceding paper,[10] it was shown that a much wider class of second-order, two-dimensional, hyperbolic differential equations, including Eqs. (5.16)—(5.18) as special cases, admits nonspreading solutions. From a comparison of Eqs. (5.6) and these other equations, it

is seen that for some Friedmann backgrounds that do not satisfy Eqs. (5.27)—(5.29), Eqs. (5.6) will nevertheless admit nonspreading solutions. We leave the discussion of these other backgrounds for future study.

C. Equations (5.8)

In order to expedite the treatment of this case, we shall make use of the relationship between Eqs. (5.6) and (5.8) that was discussed at the beginning of this section. First, however, we shall examine the behavior of the equation of state (5.19) under the background transformation $a\to 1/a$. We begin by defining

$$P^*\equiv a^2 P,\quad \rho^*\equiv a^2\rho.\qquad (5.30)$$

We then define $\tilde{P}^*$ and $\tilde{\rho}^*$ as the images of P^* and ρ^*, respectively, under the transformation $a\to 1/a$, i.e., we have

$$\tilde{P}^*\equiv P^*(1/a),\quad \tilde{\rho}^*\equiv\rho^*(1/a).\qquad (5.31)$$

It then follows, from a straightforward calculation based on Eqs. (B1) and (B2), that

$$\tilde{P}^*=-(P^*+\tfrac{2}{3}\rho^*),\quad \tilde{\rho}^*=\rho^*.\qquad (5.32)$$

We see, from Eqs. (5.30)—(5.32), that the equation of state $P=\alpha\rho$ leads to

$$\tilde{P}=\tilde{\alpha}\tilde{\rho},\qquad (5.33)$$

where

$$\tilde{P}\equiv P(1/a),\quad \tilde{\rho}\equiv\rho(1/a),\qquad (5.34)$$

and

$$\tilde{\alpha}\equiv-\alpha-\tfrac{2}{3}.\qquad (5.35)$$

In other words, if a background characterized by $a(t)$ has its pressure proportional to its density, then so does a background characterized by $1/a(t)$, with the proportionality constants for the two backgrounds related by Eq. (5.35).

Based on Eq. (5.35) and the relationship between Eqs. (5.6) and (5.8), as explained at the beginning of this section, a simple calculation shows that Eqs. (5.24) and (5.26), pertaining to Eqs. (5.6), should be replaced by

$$d=\mp\frac{4}{3\alpha+1},\qquad (5.36)$$

and

$$\alpha=-\frac{l'+3}{3(l'+1)},\quad l'=0,1,2,\cdots,\qquad (5.37a)$$

$$\alpha=\frac{2-l'}{3l'},\quad l'=1,2,\cdots,\qquad (5.37b)$$

respectively, in the present case. The numerical results of Eqs. (5.37) are indicated in Table III.

A comparison of Eqs. (5.26) and (5.37) (or of Tables II and III) shows that the set of values of α determined by Eqs. (5.26) is identical to that determined by Eqs. (5.37). Furthermore, a one-step clockwise rotation of

TABLE III. Values of α satisfying Eqs. (5.37) for allowed values of l'.

$l' = 0$	$l' = 1$	$l' = 2$	$l' = 3$	---	$l' \to \infty$
	$\alpha = \frac{1}{3}$	$\alpha = 0$	$\alpha = -\frac{1}{5}$	---	
$\alpha = -1$					$\alpha \to -\frac{1}{3}$
	$\alpha = -\frac{2}{3}$	$\alpha = -\frac{5}{9}$	$\alpha = -\frac{1}{2}$	---	

the α's in Table II (but keeping the limiting value of $-\frac{1}{3}$ in the same position) gives us Table III.

Again because of the relationship between Eqs. (5.6) and (5.8), the search for more general Friedmann backgrounds for which Eqs. (5.8) admit nonspreading solutions is essentially the same as in the case of Eqs. (5.6); we shall pursue it no further here.

6. DISCUSSION

Perhaps the most interesting of our results is that there is a class of Friedmann backgrounds, including the physically important cases $P = 0$ and $P = \rho/3$, in which gravitational and Klein—Gordon multipole radiation can propagate without backscattering. This should be contrasted with the earlier results of Kundt and Newman,[1] which suggested that the presence of matter would lead to noncharacteristic propagation.

In Sec. 5, we noted several rather mysterious results. There was a strange relation between the equations describing the Klein—Gordon and the gravitational fields; although they arose from quite different considerations, they were images of one another under the simple background transformation $a(t) \to 1/a(t)$. This transformation had the further strange consequence that it preserved a proportionality between the pressure and the density of the backgrounds. The simple relation between these constants of proportionality led to still another strange relationship, that between Tables II and III. We suspect that these results are not just mathematical coincidences, but may arise from some deeper physical structure that we have not yet understood.

Finally, it may be of some interest to pursue similar discussions of radiation in other sorts of cosmological backgrounds. A preliminary study of Klein—Gordon fields in de Sitter backgrounds indicates that multipole radiation does propagate characteristically.

APPENDIX A

In this appendix, we give a brief review of the Friedmann metrics.

It can be shown that the line element for any Friedmann model can be written in one of the following forms[11]:

$$ds^2 = a^2(t)[dt^2 - d\chi^2 - \sinh^2\chi(d\theta^2 + \sin^2\theta\, d\phi^2)], \quad \text{(A1)}$$

$$ds^2 = a^2(t)[dt^2 - d\chi^2 \Psi_2 - \sin^2\chi(d\theta^2 + \sin^2\theta\, d\phi^2)], \quad \text{(A2)}$$

$$ds^2 = a^2(t)[dt^2 - d\chi^2 - \chi^2(d\theta^2 + \sin^2\theta\, d\phi^2)]. \quad \text{(A3)}$$

Models with negative, positive, and zero spatial curvature correspond to the forms (A1), (A2), and (A3), respectively. We take $\chi > 0$ in (A1) and (A3), $0 < \chi < \pi$ in (A2), and $a > 0$ in all three cases.

If we introduce the coordinates $u = t - \chi$ and $\zeta = x^2 + ix^3$, where x^2 and x^3 are the stereographic coordinates related to θ and ϕ, then Eqs. (A1)—(A3) take the forms

$$ds^2 = a^2(t)[- du^2 + 2du\, dt - (1 + \tfrac{1}{4}\zeta\bar\zeta)^{-2} \sinh^2(t - u)\, d\zeta\, d\bar\zeta], \quad \text{(A4)}$$

$$ds^2 = a^2(t)[- du^2 + 2du\, dt - (1 + \tfrac{1}{4}\zeta\bar\zeta)^{-2} \sin^2(t - u)\, d\zeta\, d\bar\zeta], \quad \text{(A5)}$$

$$ds^2 = a^2(t)[- du^2 + 2du\, dt - (1 + \tfrac{1}{4}\zeta\bar\zeta)^{-2}(t - u)^2\, d\zeta\, d\bar\zeta], \quad \text{(A6)}$$

respectively. If we now introduce the coordinates v and v', defined by

$$v = u, \quad v' = u - 2t, \quad \text{(A7)}$$

then Eqs. (A4)—(A6) become

$$ds^2 = a^2\left(\frac{v - v'}{2}\right)\left[- dv\, dv' - (1 + \tfrac{1}{4}\zeta\bar\zeta)^{-2} \sinh^2\left(\frac{v + v'}{2}\right) d\zeta\, d\bar\zeta\right], \quad \text{(A8)}$$

$$ds^2 = a^2\left(\frac{v - v'}{2}\right)\left[- dv\, dv' - (1 + \tfrac{1}{4}\zeta\bar\zeta)^{-2} \sin^2\left(\frac{v + v'}{2}\right) d\zeta\, d\bar\zeta\right], \quad \text{(A9)}$$

$$ds^2 = a^2\left(\frac{v - v'}{2}\right)\left[- dv\, dv' - (1 + \tfrac{1}{4}\zeta\bar\zeta)^{-2}\left(\frac{v + v'}{2}\right)^2 d\zeta\, d\bar\zeta\right], \quad \text{(A10)}$$

respectively. It is clear from Eqs. (A8)—(A10) that v and v' label null hypersurfaces.

It will be useful to define the functions H and h by the equations

$$e^H \equiv (1 + \tfrac{1}{4}\zeta\bar\zeta)(h)^{1/2}, \quad \text{(A11)}$$

$$h \equiv \begin{cases} (1 + \tfrac{1}{4}\zeta\bar\zeta)^{-2}a^2 \sinh^2(t - u), \\ (1 + \tfrac{1}{4}\zeta\bar\zeta)^{-2}a^2 \sin^2(t - u), \\ (1 + \tfrac{1}{4}\zeta\bar\zeta)^{-2} a^2(t - u)^2, \end{cases} \quad \text{(A12)}$$

for models with negative, positive, and zero spatial curvature, respectively.

In our discussion of radiation, we need the zero-order values of the quantities that occur in the Newman—Penrose formalism,[5] i.e., the values associated with the Friedmann backgrounds. If we use Hawking's[9] notation and definitions of field variables, as well as his tetrad and coordinate conditions, then we find,[7] where the subscript 0 indicates that these are the zero-order quantities.

$$\underset{0}{U} = a^2/2, \quad \underset{0}{\xi^2} = i\underset{0}{\xi^3} = (2h)^{-1/2}, \quad \underset{0}{X^2} = \underset{0}{X^3} = 0, \quad \text{(A13)}$$

$$\underset{0}{\pi} = \underset{0}{\sigma} = \underset{0}{\lambda} = \underset{0}{\tau} = \underset{0}{\nu} = 0, \quad \text{(A14)}$$

$$\underset{0}{\gamma} = - a'/2a, \quad \text{(A15)}$$

$$\underset{0}{\mu} = \frac{\partial}{\partial u} \ln h^{1/2} + \frac{1}{2}\frac{\partial}{\partial t} \ln h^{1/2}, \quad \text{(A16)}$$

$$\underset{0}{\rho} = - \frac{1}{a^2}\frac{\partial}{\partial t} \ln h^{1/2} = \frac{2}{a^2}(\underset{0}{\mu} + 2\underset{0}{\gamma}), \quad \text{(A17)}$$

$$\underset{0}{\alpha} = - \underset{0}{\beta} = \frac{\zeta}{4(2h)^{1/2}(1 + \tfrac{1}{4}\zeta\bar\zeta)}, \quad \text{(A18)}$$

$$\underset{0}{\Psi_0} = \underset{0}{\Psi_1} = \underset{0}{\Psi_2} = \underset{0}{\Psi_3} = \underset{0}{\Psi_4} = 0, \quad \text{(A19)}$$

$$\underset{0}{\Phi_{01}} = \underset{0}{\Phi_{02}} = \underset{0}{\Phi_{12}} = 0, \quad \text{(A20)}$$

$$\Lambda = \begin{cases} \dfrac{a''}{4a^3} - \dfrac{1}{4a^2}\,, \\[2mm] \dfrac{a''}{4a^3} + \dfrac{1}{4a^2}\,, \\[2mm] \dfrac{a''}{4a^3}\,, \end{cases} \tag{A21}$$

$$\Phi_{00} = \frac{4}{a^2}\,\Phi_{11} = \frac{4}{a^4}\,\Phi_{22} = \begin{cases} \dfrac{2(a')^2}{a^6} - \dfrac{a''}{a^5} - \dfrac{1}{a^4}\,, \\[2mm] \dfrac{2(a')^2}{a^6} - \dfrac{a''}{a^5} + \dfrac{1}{a^4}\,, \\[2mm] \dfrac{2(a')^2}{a^6} - \dfrac{a''}{a^5}\,. \end{cases} \tag{A22}$$

In the above equations, a prime indicates differentiation with respect to t, and in Eqs. (A21) and (A22) the top, middle, and bottom lines are for models with negative, positive, and zero spatial curvature, respectively.

Finally, it will also be useful to define the function I by

$$I \equiv a^2 e^{-2H} = \begin{cases} \operatorname{csch}^2[(v + v')/2], \\[2mm] \csc^2[(v + v')/2], \\[2mm] 4/(v + v')^2, \end{cases} \tag{A23}$$

for models with negative, positive, and zero spatial curvature, respectively.

APPENDIX B

In this appendix, we study Friedmann backgrounds that are subject to the equation of state (5.19).

The pressure, P, and density, ρ, of Friedmann backgrounds can be expressed as[7]

$$P = \begin{cases} a^{-4}[(a')^2 - 2aa'' + a^2], \\[2mm] a^{-4}[(a')^2 - 2aa'' - a^2], \\[2mm] a^{-4}[(a')^2 - 2aa''], \end{cases} \tag{B1}$$

$$\rho = \begin{cases} 3a^{-4}[(a')^2 - a^2], \\[2mm] 3a^{-4}[(a')^2 + a^2], \\[2mm] 3a^{-4}(a')^2, \end{cases} \tag{B2}$$

for models with negative, positive, and zero spatial curvature, respectively. If we use Eqs. (B1) and (B2) in Eq. (5.19), we obtain

$$[a^{(3\alpha-1)/2}a']' - [(3\alpha + 1)/2]a^{(3\alpha+1)/2} = 0, \tag{B3}$$

$$[a^{(3\alpha-1)/2}a']' + [(3\alpha + 1)/2]a^{(3\alpha+1)/2} = 0, \tag{B4}$$

$$[a^{(3\alpha-1)/2}a']' = 0, \tag{B5}$$

for models with negative, positive, and zero spatial curvature, respectively.

If $3\alpha + 1 \neq 0$, we define $Y \equiv a^{(3\alpha+1)/2}$. Then Eqs. (B3)–(B5) take the forms

$$Y'' - \left(\frac{3\alpha + 1}{2}\right)^2 Y = 0, \tag{B6}$$

$$Y'' + \left(\frac{3\alpha + 1}{2}\right)^2 Y = 0, \tag{B7}$$

$$Y'' = 0, \tag{B8}$$

respectively. The general solutions to Eqs. (B6)–(B8) are

$$Y = A \sinh\left(\frac{3\alpha + 1}{2}\,t + B\right), \tag{B9}$$

$$Y = A \sin\left(\frac{3\alpha + 1}{2}\,t + B\right), \tag{B10}$$

$$Y = A(t + B), \tag{B11}$$

respectively, where A and B are arbitrary constants. We can always make $B = 0$ by an appropriate time translation. With this choice, we find

$$a = \begin{cases} \left| A \cdot \left| \sinh\left(\dfrac{3\alpha + 1}{2}\,t\right) \right| \right|^{2/(3\alpha+1)}, \\[3mm] \left| A \cdot \left| \sin\left(\dfrac{3\alpha + 1}{2}\,t\right) \right| \right|^{2/(3\alpha+1)}, \\[3mm] |A| \cdot |t|^{2/(3\alpha+1)}, \end{cases} \tag{B12}$$

for models with negative, positive, and zero spatial curvature, respectively. Equations (B12) and (5.10) immediately lead to Eqs. (5.20) for $a^2\Lambda_0$.

Finally, if $3\alpha + 1 = 0$, Eqs. (B3)–(B5) all reduce to the form

$$(\ln a)'' = 0. \tag{B13}$$

Again, one of the constants in the general solution to Eq. (B13) may be eliminated by a time translation, leaving us with the solution

$$a = e^{At}, \tag{B14}$$

where A is an arbitrary constant. Equations (B14) and (5.10) then give us

$$a^2\Lambda_0 = \begin{cases} (A^2 - 1)/4, \\[2mm] (A^2 + 1)/4, \\[2mm] A^2/4, \end{cases} \tag{B15}$$

for models with negative, positive, and zero spatial curvature, respectively.

†This work incorporates some results of a Ph.D. dissertation by S.C. Chang (University of Pittsburgh, 1974).
‡Present address: Aerojet Nuclear Company, 550 Second St., Idaho Falls, Idaho 83401.

[1] W. Kundt and E.T. Newman, J. Math. Phys. **9**, 2193 (1968).
[2] R. Penrose, in *Relativity, Groups and Topology*, edited by C. DeWitt and B. DeWitt (Gordon and Breach, New York, 1964), p. 565.
[3] L. Infeld and A. Schild, Phys. Rev. **68**, 250 (1945).
[4] L. Parker, Phys. Rev. D **5**, 2905 (1972).
[5] E.T. Newman and R. Penrose, J. Math. Phys. **3**, 566 (1962).
[6] G.E. Tauber, J. Math. Phys. **8**, 118 (1967).
[7] S.C. Chang, *Radiation in Cosmological Backgrounds*, Ph.D. dissertation (University of Pittsburgh, Pittsburgh, 1974).
[8] E.T. Newman and R. Penrose, J. Math. Phys. **7**, 863 (1966).
[9] S.W. Hawking, J. Math. Phys. **9**, 598 (1968). Equation (3.41) of this paper should be corrected to read
$$D\Phi_{22} - \delta\Phi_{21} - \bar{\delta}\Phi_{12} + \Delta\Phi_{11} + 3\Delta\Lambda = \nu\Phi_{01} + \bar{\nu}\Phi_{10} - 4\mu\Phi_{11} - \lambda\Phi_{02}$$
$$- \bar{\lambda}\Phi_{20} + (\bar{\tau} + 2\bar{\beta})\Phi_{12} + (\tau + 2\beta)\Phi_{21} + 2\rho\Phi_{22}.$$
[10] S.C. Chang and A.I. Janis, J. Math. Phys. **17**, 1429 (1976).
[11] See, for example, Eqs. (4.2), (4.21), (4.22), and (4.23) of Ref. 3.

Symmetry breaking interactions for the time dependent Schrödinger equation

C. P. Boyer,* R. T. Sharp,† and P. Winternitz

Centre de Recherches Mathématiques, Université de Montréal, Montréal, Québec, Canada

A systematic study of the symmetry porperties of the Schrödinger equation $u_{xx} + iu_t = F(x,t,u,u^*)$ is performed. The free particle equation (for $F = 0$) is known to be invariant under the six-dimensional Schrödinger group S_1. In this paper we find all continuous subgroups of S_1 and for each subgroup we construct the most general interaction term $F(x,t,u,u^*)$, reducing the symmetry group of the equation from S_1 to the considered subgroup. Since we allow for an arbitrary dependence of F on the wavefunction u (and its complex conjugate u^*) the considered Schrödinger equation is in general a nonlinear one [the ordinary Schrödinger equation with a time dependent potential is recovered if $F(x,t,u,u^*) = uG(x,t)$]. For each symmetry breaking interaction F the remaining symmetry group is used to obtain special solutions of the equations or at least to separate variables in the equation and to obtain some properties of the solutions.

1. INTRODUCTION

The purpose of this article is to provide a systematic study of the group theoretical properties of the time dependent Schrödinger equation

$$\frac{\partial^2 u}{\partial x^2} + i\frac{\partial u}{\partial t} = F(x, t, u, u^*). \qquad (1)$$

Here x and t are the space and time coordinates, u is the wavefunction, a star denotes complex conjugation, and $F(x, t, u, u^*)$ an interaction term which may depend linearly or nonlinearly on the wavefunction (we assume the absence of couplings involving derivatives of the wavefunction).

The symmetry group of the free Schrödinger equation [i.e., equation (1) with $F(x, t, u, u^*) = 0$] is known to be a six-parameter Lie group,[1-5] sometimes called the Schrödinger group, and we shall denote it S_1. We make use of recently developed methods[6-11] to find all continuous subgroups of S_1 and their invariants.

For each subgroup S_i of S_1 we then find the most general interaction of the type considered in Eq. (1) that reduces the symmetry of the system from S_1 to S_i. The subgroup S_i can be used to investigate Eq. (1) with the appropriate interaction F, in particular to separate variables in some specific system of coordinates and in some cases to obtain general or special solutions of (1).

Obviously, if the interaction is of the form

$$\Phi(x, t, u, u^*) = uV(x, t), \qquad (2)$$

then $V(x, t)$ can in general be interpreted as a time dependent potential. More generally, we obtain nonlinear interactions, e.g., the "nonlinear Schrödinger equation" $u_{xx} + iu_t = \lambda u |u|^2$, which is in itself of considerable interest.[12] Thus, the classification of subgroups also provides a method for generating soluble or at least partially soluble models with nontrivial interactions, in particular nonlinear ones. In view of the recently increased interest in nonlinear problems in physics, specially in connection with soliton type phenomena,[13-16] we find development of group theoretical methods applicable to nonlinear problems particularly opportune.

Let us put the considerations of this article into a somewhat broader mathematical and physical context. The application of Lie groups to study and solve ordinary and partial differential equations has a long history, going back to the classical work of Lie himself.[17] More recently several books have been partially or completely devoted to this subject.[18-21] Symmetry methods are very useful for treating ordinary differential equations, but their full power manifests itself best for partial differential equations. Thus, Lie theory provides a general treatment of the separation of variables in partial differential equations.[22-30] Indeed, separable systems of coordinates for a given linear partial differential equation can be characterized by the fact that the separated solutions are eigenfunctions of certain first or second order operators in the enveloping algebra of the Lie algebra of the equation's symmetry group. There is thus a correspondence between orbits of such operators under the symmetry group and different types of separable coordinates. This correspondence makes it possible to extend greatly the classes of special functions that can be treated by Lie group methods (the fact that essentailly all properties of most special functions follow directly from the representation theory of Lie groups is, of course, well known[30-33]). For both partial and ordinary differential equations Lie theory provides methods for constructing special solutions, for generating classes of solutions from one known solution, for decreasing the order of the equation or the number of variables, and generally simplifying the equation that we wish to solve. On the other hand, Lie theory makes it possible to establish relations between different equations and their solutions. Indeed, given a differential equation, we can find its symmetry group and then construct other equations, e.g., in quite different spaces, left invariant by the same group.

From the physical point of view we have the following situation. A physical system, described, e.g., by a differential equation, may have a certain symmetry, described by the symmetry group of the equation. Aside from providing methods of solving the equation, group theory allows us to approach systematically important physical problems, in particular those related to sym-

metry breaking. Thus the physical system can be placed in an external field, previously ignored interactions can be taken into account, the system can be placed in an environment that imposes certain boundary conditions, etc. All of these types of perturbing influences can be classified with respect to their symmetry, by classifying the subgroups of the original symmetry group.

Returning to the Schrödinger group S_1, or more generally S_n, as the symmetry group of the time dependent Schrödinger equation in n spatial dimensions, let us mention that it and its subgroups are also of interest in some other connections. Indeed S_n is a subgroup of the conformal group $O(n+2, 2)$ of $(n+1)$-dimensional Minkowski space; in particular S_1 is a subgroup of the de Sitter group $O(3, 2)$ and S_2 is a subgroup of the conformal group $O(4, 2)$ of Minkowski space. This plays a role in light cone or infinite momentum frame calculations in high energy physics.[34] We also mention that $O(n+2, 2)$ is the symmetry group of the Hamilton Jacobi equation in n spatial dimensions and that S_n is then related to a quantization prescription.[35]

In Sec. 2 we discuss the group S_n as the symmetry group of the free time dependent Schrödinger equation in n spatial dimensions. We explicitly construct the generators of S_1 as first order differential operators in x and t and consider some relevant properties of the Schrödinger group and its Lie algebra. All subalgebras of the Lie algebra LS_1 of S_1 are classified into conjugacy classes in Sec. 3. Conjugacy is considered separately under the subgroup D of S_1 (the Galilei group extended by dilations) and under the group S_1. The results are summarized in Table I and Fig. 2. In Sec. 4 we find for each subgroup of S_1 the most general interaction $F(x, t, u, u^*)$, left invariant by the subgroup. The results are summarized in Tables II–V. The conclusions and future outlook are presented in Sec. 5.

2. THE SCHRÖDINGER GROUP

A general method for determining the symmetry group of a given differential equation or system of differential equations is due to Lie.[17] His method is applicable for general linear or nonlinear differential equations of arbitrary order. For linear equations his method can be simplified and cast into an operator form. The equation that we wish to study in this section is indeed linear, namely the one-dimensional time dependent Schrödinger equation for a free particle. Its symmetry group is well known,[1–5] and we shall give only a sketch of the derivation here. Since the number of spatial dimensions is not crucial for the derivation, we treat the n-dimensional case and only later specify that $n = 1$. The equation to consider is thus

$$\Delta u = u_{\mathbf{xx}} + i u_t = 0, \tag{3}$$

where $u_t \equiv \partial u / \partial t$ and $u_{\mathbf{xx}} = \sum_{i=1}^{n} \partial^2 u / \partial x_i^2$ is the n-dimensional Laplace operator in the Euclidean space E_n. The operator Δ acts in a space of functions $\psi(\mathbf{x}, t)$ (to be specified below) and the group G will be called its symmetry group, if their exists a representation $T(g)$ of G, acting in the space of functions $\psi(\mathbf{x}, t)$ in such a manner that it transforms solutions of (3) into solutions. Without loss of generality we can assume that the action of $T(g)$ is

$$[T(g)\psi](\mathbf{x}, t) = \mu(\mathbf{x}, t, g)\psi(\mathbf{x}', t') + \phi(\mathbf{x}, t). \tag{4}$$

Here $\mu(\mathbf{x}, t, g)$ is a scalar multiplier and the coordinates $\mathbf{x}'$, t' refer to the transformed point: $(\mathbf{x}', t') = (\mathbf{x}, t) \cdot g$. The inhomogeneous term $\phi(\mathbf{x}, t)$ simply takes into account the superposition principle for a linear equation and will be dropped in the rest of this article. However, some remnants of this symmetry for nonlinear interactions could lead to quite nontrivial results (not considered in this article).

For G to be a symmetry group we require that

$$\Delta[T(g)u](\mathbf{x}, t) = 0 \tag{5}$$

for all $u(\mathbf{x}, t)$ satisfying (3), where

$$[T(g)u](\mathbf{x}, t) = \mu(\mathbf{x}, t, g)u(\mathbf{x}', t'). \tag{6}$$

We can assume that $u(\mathbf{x}, t)$ is infinitely differentiable in the underlying variables $(\mathbf{x}, t)$ and then make use of an infinitesimal approach. Thus, we expand the operator $T(g)$ into a power series

$$T(g) = 1 + \epsilon X(\mathbf{x}, t) + \cdots, \tag{7}$$

where X is a first order linear differential operator of the form

$$X(\mathbf{x}, t) = a(\mathbf{x}, t)\partial_t + \mathbf{b}(\mathbf{x}, t)\partial_{\mathbf{x}} + c(\mathbf{x}, t) \tag{8}$$

and a, $\mathbf{b}$ and c are functions that are twice differentiable in x^i and once in t. Equation (5) now implies the operator equation

$$[\Delta, X] = \lambda(\mathbf{x}, t)\Delta. \tag{9}$$

Since Δ is a second order operator and X a first order one, $\lambda(\mathbf{x}, t)$ can only be a function (not an operator). Inserting (8) into (9) and equating the coefficients multiplying $\partial_{x_i x_j}$, ∂_{x_i}, ∂_t, and 1, we obtain a system of differential equations for a, b, c, and λ. These can be solved[1–5] to obtain the Lie algebra of the Schrödinger group S_n for the n-dimensional Euclidean space. The structure of the group is $S_n \approx [SL(2, R) \otimes O(n)] \square W_n$, where $\otimes$ denotes a direct product and $\square$ a semidirect one, with W_n as an invariant subgroup. Here $SL(2, R)$, $O(n)$, and W_n denote the real special linear group in two dimensions, the real orthogonal and real Weyl groups in n dimensions, respectively.

We now restrict ourselves to the case $n = 1$ (one spatial dimension) where we have $S_1 \equiv S \approx SL(2, R) \square W_1$. The generators of this group in the considered (reducible) representation (6) can be written as

$$X_1 \equiv H \equiv K_2 + L_3 = \partial_t,$$

$$X_2 \equiv D \equiv 2K_1 = 2t\partial_t + x\partial_x + \tfrac{1}{2},$$

$$X_3 \equiv C \equiv -K_2 + L_3 = t^2\partial_t + tx\partial_x + t/2 - ix^2/4, \tag{10}$$

$$X_4 \equiv P = \partial_x,$$

$$X_5 \equiv B = -t\partial_x + ix/2,$$

$$X_6 \equiv E = i.$$

We consider these operators to be acting on the space $C_0^\infty(R^2)$ of infinitely differentiable functions of compact support in x and t. However, since the intersection of $C_0^\infty(R^2)$ with the null space S of Δ [solutions of the Schrödinger equation (3)] is an invariant subspace of

$C_0^\infty(R^2)$, we can restrict the generators (10) to the space $F \equiv S \cap C_0^\infty(R^2)$ and consider the time t as a parameter. Notice that on S we have $\partial_t = i\partial_{xx}$. The generators (10) have been constructed as skew-symmetric operators on F endowed with the usual quantum mechanical inner product

$$(\psi_1, \psi_2) = \int_{-\infty}^\infty dx\, \psi_1^*(x, t)\psi_2(x, t) \tag{11}$$

for $\psi_1, \psi_2 \in F$. Moreover, all the generators can be seen to be essentially skew-adjoint on F.

The above construction of a Hilbert space is quite natural from the usual quantum mechanical viewpoint. It is, however, in no sense obligatory, and in particular the structure may be destroyed by introducing nonlinear interactions, as in Sec. 4. If we drop the requirement that the underlying coordinates (x, t) remain real, we find that the generator X_2 is modified to

$$\widetilde{X}_2 = \widetilde{D} = 2t\partial_t + x\partial_x + \tfrac{1}{2} + \rho, \quad \rho = \text{real},$$

and the central element X_6 is replaced by an arbitrary complex number, i.e., the group is extended to a seven-dimensional group, generated by $X_1, \widetilde{X}_2, X_3, \ldots, X_6$, and $X_7 = 1$. We shall however restrict ourselves to the Schrödinger group $\mathcal{S}_1$, generated by (10).

The nonzero commutation relations of the generators are

$$[H, D] = 2H, \quad [H, C] = D, \quad [D, C] = 2C, \quad [P, B] = \tfrac{1}{2}E,$$
$$[H, B] = -P, \quad [D, P] = -P, \quad [D, B] = B, \quad [C, P] = B. \tag{12}$$

Thus, H, D, and C generate $SL(2, R)$, P, B, and E generate the Weyl group W the action of the algebra of $SL(2, R)$ on that of W is as given in (12). The action of the differential operators (10) can be exponentiated to give a representation[1-5,26] of the Schrödinger group $\mathcal{S}_1$. The action of the Weyl group is given by the operators

$$T(w, z, \alpha) = e^{wB}e^{zP}e^{\alpha E} \tag{13}$$

acting as

$$[T(w, z, \alpha)\psi](x, t) = \exp[(i/4)(2wx - w^2 t + 4\alpha)$$
$$\times \psi(x - wt + z, t)]. \tag{14}$$

The action of the $SL(2, R)$ subgroup is

$$[T(A)\psi](x, t) = \exp\left(\frac{ib}{4}\frac{x^2}{bt + d}\right)(d + bt)^{-1/2}$$
$$\times \psi\left(\frac{x}{d + bt}, \frac{c + at}{d + bt}\right) \tag{15}$$

with

$$A = \begin{pmatrix} a\,b \\ c\,d \end{pmatrix} \in SL(2, R), \quad ad - bc = 1.$$

The adjoint action of $SL(2, R)$ on W_1 is

$$T^{-1}(A)T(w, z, \alpha)T(A) = T(w', z', \alpha'), \tag{16}$$

where

$$w' = dw + bz, \quad z' = az + cw, \quad \alpha' = \alpha + \tfrac{1}{4}(wz - w'z'). \tag{17}$$

Note that we have

$$T\begin{pmatrix} 1 & b \\ 0 & 1 \end{pmatrix} = \exp[b(K_2 - L_3)], \quad T\begin{pmatrix} 1 & 0 \\ c & 1 \end{pmatrix} = \exp[c(K_2 + L_3)]$$

$$T\begin{pmatrix} e^a & 0 \\ 0 & e^{-a} \end{pmatrix} = \exp(aK_3), \quad T\begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix} = \exp(\theta L_3). \tag{18}$$

A comparison of (10) with (13)—(18) clarifies the notations of (10). Indeed, H (the Hamiltonian) generates time translations, D dilations, C conformal transformations, P space translations, B Galilei boosts, and the central element E corresponds to the identity transformation (and to a constant phase factor multiplying the wavefunction).

Let us make some comments on the representation (13)—(18):

(1) In view of the square root factor in (15) we have a representation of the two fold covering group of $SL(2, R)$, rather than of $SL(2, R)$ itself.

(2) The representation is not irreducible but becomes so when restricted to the space S of solutions of Eq. (3). By the extension of the restricted representation to the Hilbert space given by Eq. (11) the representation is unitary.

(3) The representation of the Schrödinger group $\mathcal{S}_1$ is irreducible on S; however, the representation (15) of the covering group of $SL(2, R)$ is the direct sum of two unitary irreducible representations, namely $D_{1/4}^+ \oplus D_{3/4}^+$ in Bargmann's notations.[36]

(4) The Schrödinger group $\mathcal{S}_1$ has two invariant operators (Casimir operators). They are

$$C^{(1)} = E,$$
$$C^{(2)} = \{4DBP + 4B^2H + 4CP^2 - ED^2 + 4ECH\}_{\text{sym}}, \tag{19}$$

where each term in $C^{(2)}$ must be symmetrized with respect to permutations of all entries, e.g., $\{DBP\}_{\text{sym}} = \tfrac{1}{6}\{DBP + DPB + BPD + BDP + PBD + PDB\}$. In the representation restricted to the subspace of solutions of (3) both of these invariants have a definite value, namely

$$C^{(1)} = i, \quad C^{(2)} = i/4. \tag{20}$$

(5) The continuous group $\mathcal{S}_1$ can be extended by the discrete transformations T, X, and XT, related to time reversal and space reflection in the underlying space (x, t). Indeed, if we consider the reversal $t \to -t$, $x \to x$ together with the complex conjugation operator K, we obtain the operator T. The operator X corresponds to the transformation $t \to t$, $x \to -x$. Clearly X, T, and XT are symmetries of Eq. (3) and a glance at the generators (10) shows the following behavior under these transformations:

$$T: K_1 \to K_1, \quad K_2 \to -K_2, \quad L_3 \to -L_3, \quad P \to P \quad B \to -B, \quad E \to -E,$$

$$X: K_1 \to K_1, \quad K_1 \to K_2, \quad L_3 \to L_3, \quad P \to -P \quad B \to -B, \quad E \to E,$$

$$XT: K_1 \to K_1, \quad K_2 \to -K_2, \quad L_3 \to -L_3, \quad P \to -P, \quad B \to B, \quad E \to -E, \tag{21}$$

3. SUBGROUPS OF THE SCHRÖDINGER GROUP S_1 AND THEIR INVARIANTS

We consider the algebra of S_1, using the basis provided by the operators K_1, K_2, L_3, P, B, and E of (10). All conjugacy classes of subalgebras can be found using known methods.[7] We shall consider conjugacy classes of subalgebras, where conjugacy is considered with respect to the Schrödinger group S_1 on one hand and on the other hand with respect to a "geometric" subgroup D of the Schrödinger group, generated by H, D, P, B, and E. This subgroup can be interpreted as a Galilei group, extended by dilations; its transformations [see (14) and (15) with $b = 0$, $d = a^{-1}$] of the underlying (x, t) manifold are linear, as opposed to the conformal transformations, generated by C.

The algorithm for classifying subalgebras into conjugacy classes with respect to some group of automorphisms A (to be identified with S_1 or D) makes use of the fact that the Weyl algebra $LW = \{P, B, E\}$ is an ideal in the considered algebra and that the factor algebra LS_1/LW is isomorphic to $LSL(2, R)$ (if G is a Lie group, LG will be its Lie algebra). The algorithm consists of the following steps:

(1) Find all subalgebras of the factor algebra $LSL(2, R)$, i.e., construct a representative F_i for each conjugacy class of such algebras under A. For each F_i find its normalizer in A, satisfying $\mathrm{Nor}_A F_i \cdot F_i \subseteq F_i$.

(2) For each subalgebra F_i find all invariant subspaces N_{ia} in LW, that also form subalgebras. Use the normalizer of F_i in A to simplify N_{ia}. The algebraic sums $F_i + N_{ia}$ for all i and all a will provide us with a list of representatives of all splitting subalgebras of S_1.

(3) To find all nonsplitting subalgebras of S_1, consider separately a subalgebra F_i of $LSL(2, R)$ together with an invariant subalgebra N_{ia} in LW. To each generator of F_i add a linear combination of all generators of LW, not contained in N_{ia}. Use transformations belonging to W to simplify the above linear combinations and then further simplify, using the normalizer of $F_i + N_{ia}$ in $SL(2, R)$. Finally, restrict the coefficients in the linear combinations to ensure that we obtain an algebra. Running through all F_i and N_{ia} we obtain a list of representatives of all nonsplitting subalgebras of S_1.

We shall first find representatives of all conjugacy classes of subalgebras of LS_1 with respect to conjugacy under the five-dimensional group D and then show how various classes collapse into one under the entire Schrödinger group S_1.

A general element of $LSL(2, R)$ can be written as $X = aK_1 + b(K_2 + L_3) + c(K_2 - L_3)$. If $c \neq 0$, the term aK_1 can be transformed into zero by a transformation $\exp[x(K_2 + L_3)]$ with an appropriate choice of x. The transformation $\exp(yK_1)$ can then be used to transform X into K_2 (if $bc > 0$), L_3 (if $bc < 0$) or $K_2 - L_3$ (if $bc = 0$, $c \neq 0$). If $c = 0$, $a \neq 0$, then $\exp[x(K_2 + L_3)]$ can be used to cancel the $K_2 + L_3$ term, yielding the algebra K_1. If $a = c = 0$, we obtain $K_2 + L_3$. Thus, we obtain five nonconjugate one-dimensional algebras. Now consider each of the algebras K_1, K_2, L_3, $K_2 + L_3$, and $K_2 - L_3$ and add to it a second generator X. We must choose a, b, and c so

as to obtain an algebra satisfying $[A, B] = B$ $\{[A, B] = 0$ is not contained in $LSL(2, R)\}$. We obtain two algebras: $\{K_1, K_2 + L_3\}$ and $\{K_1, K_2 - L_3\}$. On Fig. 1 we present the subalgebras of $LSL(2, R)$, classified in this manner and also the well-known classification with respect to $SL(2, R)$. The trivial subalgebra $\{0\}$ should be added to both schemes of subalgebras.

For each subalgebra F_i of $SL(2, R)$ we must now go through all steps of our classification algorithm. As an illustration we do this for the subalgebra $F_2 = \{K_1, K_2 + L_3\}$. Commuting K_1 and $K_2 + L_3$ with the element $pP + bB + aE$, where p, b, and a are real numbers, we find the following subalgebras of LW, that are invariant under F_2:

$$N_{2,1} = \{P, B, E\}, \quad N_{2,2} = \{P, E\}, \quad N_{2,3} = \{P\}, \quad N_{2,4} = \{E\},$$
$$N_{2,5} = \{0\}. \tag{22}$$

(Note that $\{P, B\}$ forms an invariant subspace, not, however, an algebra). All splitting subalgebras are obtained as the algebraic sums $F_2 + N_{2,k}$ with $k = 1, \ldots, 5$. Now let us find all nonsplitting subalgebras. Write two possible generators in the form

$$\tilde{K}_1 = K_1 + b_1 B + p_1 P + a_1 E, \quad \tilde{K}_2 + \tilde{L}_3 = K_2 + L_3 + b_2 B$$
$$+ p_2 P + a_2 E. \tag{23}$$

Consider the individual invariant subspaces. If we add $\{P, B, E\}$ to (22), we obtain a splitting subalgebra. Adding $\{P, E\}$, we can put $p_1 = p_2 = a_1 = a_2 = 0$. The requirement that we obtain an algebra implies $b_1 = b_2 = 0$, so that the subalgebra again splits. The subalgebra $\{N_{2,3}\} = P$ will not in general be invariant. However, if $b_2 = 0$, then $P - b_1 E$ is invariant and the transformation $\exp(2b_1 B)$ will cancel the term $- b_1 E$. The requirement that we obtain an algebra then implies $a_2 = b_1 = b_2 = 0$, and we can also put $p_1 = p_2 = 0$. Thus we obtain a nonsplitting subalgebra: $\{K_1 + aE, K_2 + L_3, P, a \neq 0\}$. Taking the subalgebra $N_{2,4} = \{E\}$, we put $a_1 = a_2 = 0$. The transformation $\exp(2b_1 B - 2p_1 P)$ will cancel the term $b_1 B + p_1 P$ in $\tilde{K}_1$. The requirement that we obtain an algebra implies $b_2 = p_2 = 0$ and so the algebra splits. Finally, consider the trivial subalgebra $N_{2,5}$. We can turn b_1 and p_1 into zero by the transformation $\exp(2b_1 B - 2p_1 P)$. The
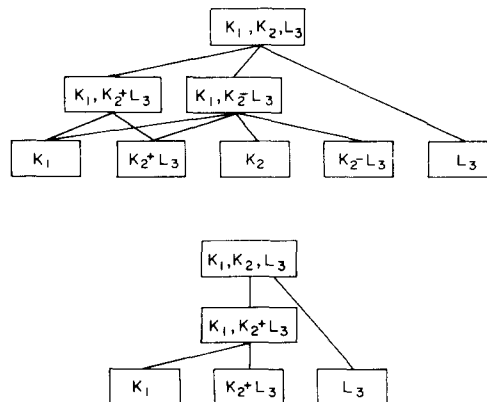

FIG. 1. (a) Subalgebras of $LSL(2,R)$ classified with respect to $C(1) \sim \exp\{K_1, K_2 + L_3\}$. (b) Subalgebras of $LSL(2,R)$ classified with respect to $SL(2,R)$.

TABLE I. Subalgebras of $L\mathcal{S}_1$ classified under the groups $\mathcal{D}$ and $\mathcal{S}_1$.

Symbol	F_i	Generators	$\dim S_{i,k}$	Type of algebra	Invariants	Comments
$S_{1,1}$	F_1	$;K_1,K_2,L_3,P,B,E$	6	$L\mathcal{S}_1$	$C^{(1)},C^{(2)}$ of (19)	
$S_{1,2}$		$E;K_1,K_2,L_3$	4	$A_1\oplus A_{3,8}$	$E,K_1^2+K_2^2-L_3^2$	
$S_{1,3}$		$;K_1,K_2,L_3$	3	$A_{3,8}$	$K_1^2+K_2^2-L_3^2$	
$S_{2,1}$	F_2	$K_1;K_2+L_3,P,B,E$	5	$L\mathcal{D}$	E	
$S_{2,2}$		$E,K_1;K_2+L_3,P$	4	$A_1\oplus A_{3,5}^h,\ h=\frac{1}{2}$	$E,(K_2+L_3)/P^2$	
$S_{2,3}$		$K_1;K_2+L_3,P$	3	$A_{3,5}^h,\ h=\frac{1}{2}$	$(K_2+L_3)/P^2$	
$S_{2,4}$		$E,K_1;K_2+L_3$	3	$A_1\oplus A_2$	E	
$S_{2,5}$		$K_1;K_2+L_3$	2	A_2	none	
$S_{2,6}$		$K_1+aE;K_2+L_3,P$ $a\neq 0$	3	$A_{3,5}^h,\ h=\frac{1}{2}$	$(K_2+L_3)/P^2$	$T\Rightarrow a>0$
$S_{2,7}$		$K_1+aE;K_2+L_3,\ a\neq 0$	2	A_2	none	$T\Rightarrow a>0$
$\widetilde{S}_{2,1}$	$\widetilde{F}_2$	$K_1;K_2-L_3,P,B,E$	5	$L\mathcal{D}$	E	$\widetilde{S}_{2,k}$ is conjugate to $S_{2,k}$
$\widetilde{S}_{2,2}$		$E,K_1;K_2-L_3$	4	$A_1\oplus A_{3,5}^h,\ h=\frac{1}{2}$	$E,(K_2-L_3)/B^2$	$(k=1,\dots,7)$ under the
$\widetilde{S}_{2,3}$		$K_1;K_2-L_3,\ B$	3	$A_{3,5}^h,\ h=\frac{1}{2}$	K_2-L_3/B^2	group $\mathcal{S}_1$
$\widetilde{S}_{2,4}$		$E,K_1;K_2-L_3$	3	$A_1\oplus A_2$	E	
$\widetilde{S}_{2,5}$		$K_1;K_2-L_3$	2	A_2	none	
$\widetilde{S}_{2,6}$		$K_1+aE;K_2-L_3,\ B$ $a\neq 0$	3	$A_{3,5}^h,\ h=\frac{1}{2}$	$(K_2-L_3)/B^2$	$T\Rightarrow a>0$
$\widetilde{S}_{2,7}$		$K_1+aE;K_2-L_3,\ a\neq 0$	2	A_2	none	$T\Rightarrow a>0$
$S_{3,1}$	F_3	$K_1;P,B,E$	4	$A_{4,8}$	$E,2K_1E+PB+BP$	
$S_{3,2}$		$E,K_1;P$	3	$A_1\oplus A_2$	E	
$S'_{3,2}$		$E,K_1;B$	3	$A_1\oplus A_2$	E	conjugate to $S_{3,2}$ under $\mathcal{S}_1$
$S_{3,3}$		$K_1;P$	2	A_2	none	
$S'_{3,3}$		$K_1;B$	2	A_2	none	conjugate $S_{3,3}$ under $\mathcal{S}_1$
$S_{3,4}$		$K_1,E;$	2	$2A_1$	K_1,E	
$S_{3,5}$		$K_1;$	1	A_1	K_1	
$S_{3,6}$		$K_1+aE;P,\ a\neq 0$	2	A_2	none	$T\Rightarrow a>0$
$S'_{3,6}$		$K_1+aE;B,\ a\neq 0$	2	A_2	none	$T\Rightarrow a>0$; conjugate to $S_{3,6}$ under $\mathcal{S}_1$
$S_{3,7}$		$K_1+aE;\ a\neq 0$	1	A_1	K_1+aE	T or $\mathcal{S}_1\Rightarrow a>0$
$\widetilde{S}_{3,1}$	$\widetilde{F}_3$	$K_2;P,B,E$	4	$A_{4,8}$	$E,2K_2E+B^2-P^2$	$\widetilde{S}_{3,k}(\widetilde{S}'_{3,k})$ is conjugate to
$\widetilde{S}_{3,2}$		$E,K_2;B+P$	3	$A_1\oplus A_2$	E	$S_{3,k}(S'_{3,k})$ under $\mathcal{S}_1$
$\widetilde{S}'_{3,2}$		$E,K_2;B-P$	3	$A_1\oplus A_2$	E	$(k=1,\dots,7)$
$\widetilde{S}_{3,3}$		$K_2;B+P$	2	A_2	none	
$\widetilde{S}'_{3,3}$		$K_2;B-P$	2	A_2	none	
$\widetilde{S}_{3,4}$		$K_2,E;$	2	$2A_1$	K_2,E	
$\widetilde{S}_{3,5}$		$K_2;$	1	A_1	K_2	
$\widetilde{S}_{3,6}$		$K_2+aE;B+P,\ a\neq 0$	2	A_2	none	$T\Rightarrow a>0$
$\widetilde{S}'_{3,6}$		$K_2+aE;B-P,\ a\neq 0$	2	A_2	none	$T\Rightarrow a>0$
$\widetilde{S}_{3,7}$		$K_2+aE;a\neq 0$	1	A_1	K_2+aE	T or $\mathcal{S}_1\Rightarrow a>0$
$S_{4,1}$	F_4	$K_2+L_3,\ B;P,E$	4	$A_{4,1}$	$E(K_2+L_3)-P^2$	
$S_{4,2}$		$K_2+L_3,\ P,E;$	3	$3A_1$	K_2+L_3,P,E	
$S_{4,3}$		$K_2+L_3,P;$	2	$2A_1$	K_2+L_3,P	
$S_{4,4}$		$K_2+L_3,E;$	2	$2A_1$	K_2+L_3,E	
$S_{4,5}$		$K_2+L_3;$	1	A_1	K_2+L_3	
$S_{4,6}$		$K_2+L_3+\epsilon B,\ P;E,$ $\epsilon=\pm1$	3	$A_{3,1}$	E	$\mathcal{S}_1$ or $X\Rightarrow\epsilon=1$
$S_{4,7}$		$K_2+L_3+\epsilon E,P;$	2	$2A_1$	$K_2+L_3+\epsilon E,P$	
$S_{4,8}$		$K_2+L_3+\epsilon B,\ E;$ $\epsilon=\pm1$	2	$2A_1$	$K_2+L_3+\epsilon B,E$	$\mathcal{S}_1$ or $X\Rightarrow\epsilon=1$
$S_{4,9}$		$K_2+L_3+\epsilon B;\epsilon=\pm1$	1	A_1	$K_2+L_3+\epsilon B$	$\mathcal{S}_1$ or $X\Rightarrow\epsilon=1$
$S_{4,10}$		$K_2+L_3+\epsilon E;\epsilon=\pm1$	1	A_1	$K_2+L_3+\epsilon E$	
$\widetilde{S}_{4,1}$	$\widetilde{F}_4$	$K_2-L_3,\ P;B,E$	4	$A_{4,1}$	$E(K_2-L_3)+B^2$	$\widetilde{S}_{4,k}$ is conjugate to
$\widetilde{S}_{4,2}$		$K_2-L_3,B,E;$	3	$3A_1$	K_2-L_3,B,E	$S_{4,k}\ (k=1,\dots,10)$
$\widetilde{S}_{4,3}$		$K_2-L_3,\ B;$	2	$2A_1$	K_2-L_3,B	under $\mathcal{S}_1$
$\widetilde{S}_{4,4}$		$K_2-L_3,\ E;$	2	$2A_1$	K_2-L_3,E	
$\widetilde{S}_{4,5}$		$K_2-L_3;$	1	A_1	K_2-L_3	
$\widetilde{S}_{4,6}$		$K_2-L_3+\epsilon P,\ B;E,$ $\epsilon=\pm1$	3	$A_{3,1}$	E	$\mathcal{S}_1$ or $X\Rightarrow\epsilon=1$
$\widetilde{S}_{4,7}$		$K_2-L_3+\epsilon E,\ B;\epsilon=\pm1;$	2	$2A_1$	$K_2-L_3+\epsilon E,B$	
$\widetilde{S}_{4,8}$		$K_2-L_3+\epsilon P,\ E;$ $\epsilon=\pm1$	2	$2A_1$	$K_2-L_3+\epsilon P,E$	$\mathcal{S}_1$ or $X\Rightarrow\epsilon=1$
$\widetilde{S}_{4,9}$		$K_2-L_3+\epsilon P;\epsilon=\pm1$	1	A_1	$K_2-L_3+\epsilon P$	$\mathcal{S}_1$ or $X\Rightarrow\epsilon=1$
$\widetilde{S}_{4,10}$		$K_2-L_3+\epsilon E,\ \epsilon=\pm1$	1	A_1	$K_2-L_3+\epsilon E$	

TABLE I. (Continued).

$S_{5,1}$	F_5	$L_3;P,B,E$	4	$A_{4,10}$	$E,-2L_3E+P^2+B^2$	
$S_{5,2}$		L_3,E;	2	$2A_1$	L_3,E	
$S_{5,3}$		L_3;	1	A_1	L_3	
$S_{5,4}$		$L_3+aE;a\neq 0$	1	A_1	L_3+aE	
$S_{6,1}$	F_6	$P,B;E$	3	$A_{3,1}$	E	
$S_{6,2}'$		B,E;	2	$2A_1$	B,E	
$\widetilde{S}_{6,2}$		P,E;	2	$2A_1$	P,E	Conjugate to $S_{6,2}$ under $\mathcal{S}_1$
$S_{6,3}$		B;	1	A_1	B	
$\widetilde{S}_{6,3}$		P;	1	A_1	P	conjugate to $S_{6,3}$ under $\mathcal{S}_1$
$S_{6,4}$		E	1	A_1	E	
$S_{6,5}$		0	0	0	0	

condition $[\widetilde{K}_1, \widetilde{K}_2 + \widetilde{L}_3] = -\widetilde{K}_2 - \widetilde{L}_3$ then implies $b_2 = p_2 = e_2$ = 0. Thus we obtain the nonsplitting subalgebra $\{K_1 + aE, K_2 + L_3, a \neq 0\}$.

Proceeding in the same manner as above, we find the representatives of all classes of subalgebras of $L\mathcal{S}_1$. The results are summarized in Table I. Each algebra in this table represents a D conjugacy class of subalgebras of $L\mathcal{S}_1$. In the first column we introduce a symbol S_{jk}, $\widetilde{S}_{jk}$, or S_{jk}' for each subalgebra. Leaving out all $\widetilde{S}_{jk}$ and S_{jk}' algebras, we obtain the shorter list of $\mathcal{S}_1$ conjugacy classes of subalgebras. The second column gives the subalgebra F_j of LSL$(2, R)$ from which $S_{j,k}$ was produced. The generators of $S_{j,k}$ are in column 3; those to the right of the semicolon are also contained in the derived algebra of $S_{j,k}$. In some cases, e.g., $S_{2,6}$, $S_{2,7}$ et al., the algebra depends on a parameter. Its range as given in column 3 refers to D conjugacy classes. For $\mathcal{S}_1$ conjugacy classes the range may be smaller (e.g., $a > 0$ rather than $a \neq 0$, or $\epsilon = 1$ instead of $\epsilon = \pm 1$). Such cases are pointed out in the last column. In the fourth column

we give the dimension of $S_{j,k}$. The type of algebra is given in the fifth column. The notations are those used in papers[9] and are related to a classification of low-dimensional real Lie algebras, due to Mubarakzyanov.[37] Thus, A_1 is a one-dimensional real Lie algebra, nA_1 denotes a direct sum of n such algebras and A_2 is a two-dimensional non-Abelian Lie algebra (with a basis satisfying $[X, Y] = X$). Three-dimensional Lie algebras are denoted $A_{3,1}, \ldots, A_{3,9}$ and a superscript, if present (e.g., $A_{3,5}^h$) indicates that the algebra itself depends on a parameter h. The four-dimensional Lie algebras are similarly denoted $A_{4,1}, \ldots, A_{4,12}$. The commutation relations for each algebra are given elsewhere,[9] and there is no need to repeat them here, since they can be read off from the commutation relations of the generators in the third column. In the sixth column we list the invariants of all subalgebras. They can be obtained using a method described elsewhere.[9] Note that for subalgebras of $L\mathcal{S}_1$ all invariants are either polynomials in the generators (Casimir operators) or rational invariants, like e.g., $(K_2 + L_3)/P^2$ for $S_{2,2}$. The meaning of such invari-
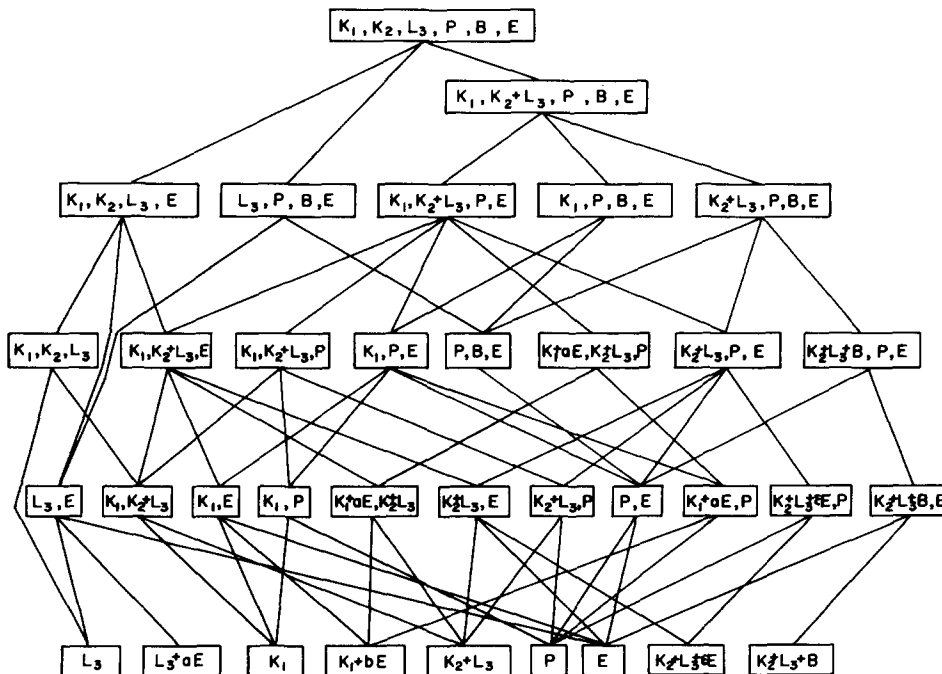
FIG. 2. Subalgebras of $L\mathcal{S}_1$ classified under $\mathcal{S}_1$ ($a \neq 0$, $b > 0$, $\epsilon = \pm 1$).

ants, as well as more general types of nonpolynomial invariants, has been discussed, e.g., in papers.[9] In the seventh column we indicate additional equivalencies between various subalgebras, when conjugacy is considered under S_1, rather than D. We also point out all cases when the inclusion of S_1 or of time reversal T leads to a further restriction on the range of the parameters a or ϵ, figuring in the third column.

On Fig. 2 we present the lattice of subalgebras of LS_1, classified under S_1. By convention we consider all parameters a to satisfy $-\infty < a < \infty$, $a \neq 0$, whereas $b > 0$ and $\epsilon = \pm 1$. Conjugacy is considered under the continuous group S_1, not including the discrete transformations T, X, or XT.

4. SYMMETRY BREAKING INTERACTIONS

A. General invariance conditions

In Sec. 2 we found the Lie group leaving Eq. (3), i.e., $\Delta u = u_{xx} + i u_t = 0$ invariant, namely the Schrödinger group S_1. In Sec. 3 we found all Lie subgroups of S_1. Here we pose a different problem, namely, for each subgroup of S_1 we wish to find the most general interaction of the form $F(x, t, u, u^*)$ such that the equation

$$\Delta u = u_{xx} + i u_t = F(x, t, u, u^*) \qquad (24)$$

is invariant under this subgroup.

Indeed, consider a one-dimensional subgroup g of S_1, transforming the space—time manifold as in

$$(x', t') = (x, t) \cdot g \qquad (25)$$

and consider the representation

$$[T_g u](x, t) = [e^{\alpha X} u](x, t) = \mu(g, x', t') u(x', t'), \qquad (26)$$

where X is the generator of T_g, α is a real parameter, and $\mu(g, x', t')$ is a multiplier. Expanding T_g into a Taylor series about the point $\alpha = 0$ [we have $g = g(\alpha)$, $g(0) = 1$], we find

$$Xu = \left\{ \frac{dx'}{d\alpha} \partial_x + \frac{dt'}{d\alpha} \partial_t + \frac{d\mu(g, x', t')}{d\alpha} \right\} u$$
$$\equiv \{ a(x, t) \partial_t + b(x, t) \partial_x + c(x, t) \} u. \qquad (27)$$

Referring back to Eq. (10), we can expand the generator X as follows:

$$X = \sum_{j=1}^{6} \alpha_j X_j, \qquad (28)$$

so that $a(x, t)$, $b(x, t)$ can be written as specific known functions of the real parameters α_j. Notice that a and b are real, c is in general complex.

The condition that Eq. (24) remains invariant under the subgroup implies

$$\Delta [T(g)u](x, t) = F(x, t, [T(g)u](x, t), [T(g)u]^*(x, t)). \qquad (29)$$

To obtain an operator formulation of invariance, we remember that $g = g(\alpha)$, take the derivative of both sides of (29) with respect to α, and then set $\alpha = 0$. We obtain

$$\Delta(Xu) = (Xu)F_u + (Xu)^* F_{u^*}. \qquad (30)$$

Since X is a subalgebra of LS_1, it must also satisfy (9), i.e.,

$$[\Delta, X] = \lambda(x, t)\Delta. \qquad (31)$$

Formulas (30) and (31) imply

$$(X + \lambda)F = (Xu)F_u + (Xu)^* F_{u^*}. \qquad (32)$$

Using (27) and remembering that F depends on x and t both explicitly and implicitly via u and u^*, we obtain a first order partial differential equation for the interaction F:

$$a(x, t)F_t + b(x, t)F_x - cuF_u - c^* u^* F_{u^*} = -(c + \lambda)F. \qquad (33)$$

To solve (33), we shall solve the subsidiary equations

$$\frac{dt}{a} = \frac{dx}{b} = -\frac{du}{cu} = -\frac{du^*}{c^* u^*} = -\frac{dF}{(c + \lambda)F}, \qquad (34)$$

and in general obtain F as some known function times an arbitrary function of three variables, obtained by solving the first three of equations (34).

Thus, for a given operator, or set of operators X, forming a subalgebra of LS_1, we find the invariant interaction $F(x, t, u, u^*)$ by solving Eq. (33) (or several such equations for higher dimensional subalgebras).

Specific solutions of the Schrödinger equation (24) can then be found that will also satisfy the equation

$$Xu = au_t + bu_x + cu = 0. \qquad (35)$$

The meaning of Eq. (35) is that the wavefunction $u(x, t)$, in addition to being a solution of the Schrödinger equation (24), is also an absolute invariant[38] of the generator X. It is precisely this additional requirement (35) that allows us to simplify (24), in particular to separate variables.

To simplify further calculations, we note that in Eqs. (33), (35) we have

$$a(x, t) \equiv a = \alpha_1 + 2\alpha_2 t + \alpha_3 t^2,$$
$$b(x, t) \equiv b = \alpha_2 x + \alpha_3 xt + \alpha_4 - \alpha_5 t,$$
$$c(x, t) \equiv c = \tfrac{1}{2}\alpha_2 + \alpha_3(2t - ix^2)/4 + \tfrac{1}{2}i\alpha_5 x + i\alpha_6, \qquad (36)$$
$$c(x, t) + \lambda(x, t) \equiv c + \lambda = \tfrac{5}{2}\alpha_2 + \alpha_3(10t - ix^2)/4$$
$$+ \tfrac{1}{2}i\alpha_5 x + i\alpha_6.$$

We now proceed to consider each subalgebra of LS_1 separately, making use of Table I.

B. One-dimensional subalgebras

We have all together 15 one-dimensional subalgebras in Table I. We shall run through all of them in this paragraph; however, K_1 and $K_1 + aE$, K_2 and $K_2 + aE$, L_3 and $L_3 + aE$, $K_2 + L_3$ and $K_2 + L_3 + \epsilon E$, and $K_2 - L_3$ and $K_2 - L_3 + \epsilon E$ will be combined together. We shall consider one of the subalgebras, namely $2L_3 + aE$ in some detail (this is one of the more complicated cases) and then only list the results for the other subalgebras.

(1) $2L_3 + aE = X_1 + X_3 + aX_6 = (1 + t^2)\partial_t + tx\partial_x + (2t - ix^2 + 4ia)/4$: We take $\alpha_1 = \alpha_3 = 1$ (all other α_i equal to zero) in (36) and write out the subsidiary equations (34):

$$\frac{dt}{1 + t^2} = \frac{dx}{xt} = -\frac{4du}{(4ia + 2t - ix^2)u} = -\frac{4du^*}{(-4ia + 2t + ix^2)u^*}$$
$$= -\frac{4dF}{(4ia + 10t - ix^2)F}. \qquad (37)$$

The first equation provides us with the "similarity variable" to be used, namely

$$\xi = x/(1+t^2)^{1/2} \tag{38}$$

(we shall always keep $\eta=t$ as the other variable except for the case when $\xi=t$ is the similarity variable; then we put $\eta=x$). From (37) we also obtain

$$u = (1+t^2)^{-1/4} \exp(i\xi^2 t/4) \left(\frac{1-it}{1+it}\right)^{a/2} \phi(\xi) \tag{39}$$

and the complex conjugate equation. By construction the function u in (39) satisfies

$$(2L_3 + aE)u = 0.$$

Finally, the subsidiary equations (37) imply that the invariant interaction has the form:

$$\begin{aligned}
F(x,t,u,u^*) &= \frac{u}{1+t^2} \widetilde{G}(\xi, \phi(\xi), \phi^*(\xi)) \\
&= \frac{u}{1+t^2} G\left(\xi, |u|^2(1+t^2)^{1/2}, \frac{u}{u^*} \exp(-i\xi^2 t/2)\right) \\
&\times \left(\frac{1+it}{1-it}\right)^a.
\end{aligned} \tag{40}$$

Substituting (39) and (40) into the Schrödinger equation $u_{xx} + iu_t = F$, we find that $\phi(\xi)$ obeys

$$\phi''(\xi) + (-\xi^2/4 + a)\phi(\xi) = \phi(\xi)G(\xi, \phi\phi^*, \phi/\phi^*). \tag{41}$$

In Eq. (41) the parameter a, also figuring in the generator $2L_3 + aE$ and in the interaction (40), plays a role analogous to that of an eigenvalue. For a general non-linear interaction (40) the value of a is fixed. However, if the interaction is linear, i.e., $G=G(\xi)$, or even if the interaction is nonlinear, but does not depend on the phase of u, i.e., $G=G(\xi, |u|^2(1+t^2)^{1/2})$, then a in (41) can be considered to be a free parameter, and we obtain different solutions of the same Schrödinger equation for different values of a. In particular, in the linear case we can thus obtain a complete set of solutions.

Thus, invariance under the one-dimensional Lie group generated by $2L_3 + aE$ leads to several typical features, namely the subsidiary equations (37), the variables $\xi = x/(1+t^2)^{1/2}$ and $\eta=t$, the function $u(x,t)$ in the form (39), satisfying (41) and the interaction (40). Let us now list the analogous features for the other one dimensional subalgebras.

(2) $2K_1 + aE = X_2 + aX_6 = 2t\partial_t + x\partial_x + \frac{1}{2} + ia,$ $-\infty<a<\infty$:

$$\frac{dt}{2t} = \frac{dx}{x} = -\frac{du}{(\frac{1}{2}+ia)u} = -\frac{du^*}{(\frac{1}{2}-ia)u^*} = -\frac{dF}{(\frac{5}{2}+ia)F}$$

$$\xi = \frac{x}{t^{1/2}}, \quad \eta=t,$$

$$F = \frac{u}{t} G\left(\xi, uu^*\sqrt{t}, \frac{u}{u^*} t^{ia} \exp\left(-\frac{i\xi^2}{4}\right)\right),$$

$$u = t^{-(1+2ia)/4} \exp\left(\frac{i\xi^2}{8}\right) \phi(\xi), \tag{42}$$

$$\phi''(\xi) + \left(\frac{\xi^2}{16} + \frac{a}{2}\right)\phi(\xi) = \phi(\xi)G\left(\xi, |\phi|^2, \frac{\phi}{\phi^*}\right) \tag{43}$$

(3) $K_2 + L_3 + \kappa E = X_1 + \kappa X_6 = \partial_t + i\kappa,$ $\kappa=0$ or $\kappa=\pm1$:

$$\frac{dt}{1} = \frac{dx}{0} = -\frac{du}{i\kappa u} = -\frac{du^*}{i\kappa u^*} = -\frac{dF}{i\kappa F}.$$

$$\xi=x, \quad \eta=t,$$

$$F = uG\left(x, uu^*, \frac{u}{u^*} \exp(2i\kappa t)\right), \quad u = \exp(-i\kappa t)\phi(x) \tag{44}$$

$$\phi''(x) + \kappa\phi(x) = \phi(x)G\left(x, |\phi|^2, \frac{\phi}{\phi^*}\right) \tag{45}$$

(4) $K_2 + L_3 + \epsilon B = X_1 + \epsilon X_5 = \partial_t + \epsilon(-t\partial_x + \frac{1}{2}ix),$ $\epsilon=\pm1$:

$$\frac{dt}{1} = \frac{dx}{-\epsilon t} = -\frac{2du}{i\epsilon xu} = \frac{2du^*}{i\epsilon xu^*} = -\frac{2dF}{i\epsilon xF},$$

$$\xi = x + \frac{1}{2}\epsilon t^2, \quad \eta=t,$$

$$F = uG\left(x + \frac{1}{2}\epsilon t^2, uu^*, \frac{u}{u^*} \exp[i(\epsilon\xi - \frac{1}{6}t^2)]\right),$$

$$u = \exp[-\frac{1}{2}it(\epsilon\xi - \frac{1}{6}t^2)]\phi(\xi), \tag{46}$$

$$\phi''(\xi) + \frac{1}{2}\epsilon\xi\phi(\xi) = \phi(\xi)G\left(\xi, \phi\phi^*, \frac{\phi}{\phi^*}\right). \tag{47}$$

(5) $P = X_4 = \partial_x$:

$$\frac{dt}{0} = \frac{dx}{1} = \frac{du}{0} = \frac{du^*}{0} = \frac{dF}{0},$$

$$\xi=t, \quad \eta=x,$$

$$F = F(t,u,u^*) = uG\left(t, |u|^2, \frac{u}{u^*}\right), \quad u=\phi(t), \tag{48}$$

$$i\phi'(t) = \phi(t)G\left(t, |\phi|^2, \frac{\phi}{\phi^*}\right). \tag{49}$$

(6) $E = X_6 = i$:

$$\frac{dt}{0} = \frac{dx}{0} = -\frac{du}{iu} = \frac{du^*}{iu^*} = -\frac{dF}{iF},$$

$$F = uG(x, t, uu^*), \quad u=u(x,t), \tag{50}$$

$$u_{xx} + iu_t = uG(x, t, |u|^2). \tag{51}$$

(7) $2K_2 + aE = X_1 - X_3 + aX_6 = (1 - t^2)\partial_t - tx\partial_x + \frac{1}{4}(-2t + ix^2 + 4ia),$ $-\infty < a < \infty$:

$$\begin{aligned}
\frac{dx}{1-t^2} &= \frac{dx}{-tx} = \frac{4du}{(2t-ix^2-4ia)u} = \frac{4du^*}{(2t+ix^2+4ia)u^*} \\
&= \frac{4dF}{(10t-ix^2-4ia)F},
\end{aligned}$$

$$\xi = \begin{cases} x/(t^2 - 1)^{1/2}, & t^2 > 1, \quad \eta=t, \\ x/(1 - t^2)^{1/2}, & t^2 < 1, \end{cases}$$

$$\begin{aligned}
F &= \frac{u}{|t^2-1|} G\left(\xi, |u|^2|t^2-1|^{1/2}, \frac{u}{u^*}\left|\frac{t+1}{t-1}\right|^{ia}\right. \\
&\left. \times \exp\left(-\frac{ix^2t}{2(t^2-1)}\right)\right)
\end{aligned} \tag{52}$$

$$u = |t^2-1|^{-1/4}\left|\frac{1-t}{1+t}\right|^{ia/2} \exp\left(\frac{ix^2t}{4(t^2-1)}\right) \phi(\xi),$$

$$\phi''(\xi) + (\frac{1}{4}\xi^2 - a)\phi = \phi G\left(\xi, |\phi|^2, \frac{\phi}{\phi^*}\right). \tag{53}$$

(8) $-K_2 + L_3 + \kappa E = X_3 + \kappa X_6 = t^2\partial_t + tx\partial_x + \frac{1}{2}t - \frac{1}{4}ix^2 + \kappa i,$

$\kappa=0$ or $\kappa=\pm1$:

$$\frac{dt}{t^2} = \frac{dx}{xt} = -\frac{4du}{(2t-ix^2+4i\kappa)u} = \frac{4du^*}{(2t+ix^2-4i\kappa)u^*}$$

$$= -\frac{4dF}{(10t - ix^2 + 4i\kappa)F},$$

$$\xi = \frac{x}{t}, \quad \eta = t,$$

$$F = \frac{u}{t^2} G\left(\xi, |u|^2 t, \frac{u}{u^*} \exp\left[-2i\left(\frac{\kappa}{t} + \frac{\xi^2 t}{4}\right)\right]\right),$$

$$u = t^{-1/2} \exp\left[i\left(\frac{\kappa}{t} + \frac{\xi^2 t}{4}\right)\right]\phi(\xi) \tag{54}$$

$$\phi''(\xi) + \kappa\phi(\xi) = \phi G\left(\xi, |\phi|^2, \frac{\phi}{\phi^*}\right). \tag{55}$$

(9) $-K_2 + L_3 + \epsilon P = X_3 + \epsilon X_4 = t^2 \partial_t + (tx + \epsilon)\partial_x + \frac{1}{4}(2t - ix^2),$

$\epsilon = \pm 1$:

$$\frac{dt}{t^2} = \frac{dx}{tx + \epsilon} = \frac{4du}{(ix^2 - 2t)u} = -\frac{4du^*}{(ix^2 + 2t)u^*} = \frac{4dF}{(ix^2 - 10t)F},$$

$$\xi = \frac{x}{t} + \frac{\epsilon}{2t^2}, \quad \eta = t,$$

$$F = \frac{u}{t^2} G\left(\xi, |u|^2 t, \frac{u}{u^*} \exp\left[-\frac{i}{2t}\left(\xi^2 t^2 + \epsilon\xi - \frac{1}{12t^2}\right)\right]\right),$$

$$u = t^{-1/2} \exp\left[\frac{i}{4t}\left(\xi^2 t^2 + \epsilon\xi - \frac{1}{12t^2}\right)\right]\phi(\xi), \tag{56}$$

$$\phi''(\xi) + \frac{1}{2}\epsilon\xi\phi(\xi) = \phi(\xi)G\left(\xi, |\phi|^2, \frac{\phi}{\phi^*}\right). \tag{57}$$

(10) $B = X_5 = -t\partial_x + \frac{1}{2}ix$:

$$\frac{dt}{0} = \frac{dx}{-t} = -\frac{2du}{ixu} = \frac{2du^*}{ixu^*} = -\frac{2dF}{ixF},$$

$$\xi = t, \quad \eta = x,$$

$$F = uG\left(t, |u|^2, \frac{u}{u^*} \exp\left(-\frac{ix^2}{2t}\right)\right), \quad u = \exp\left(\frac{ix^2}{4t}\right)\phi(t) \tag{58}$$

$$i\left(\phi'(t) + \frac{1}{2t}\phi(t)\right) = \phi(t)G\left(t, |\phi|^2, \frac{\phi}{\phi^*}\right). \tag{59}$$

Let us briefly summarize the results and make some comments.

(i) For each one-dimensional subalgebra X we have obtained an invariant interaction $F(x, t, u, u^*)$ that can be written in the form

$$F = \frac{u}{f^2(t)} G\left(\xi, |u|^2 f(t), \frac{u}{u^*} h(\xi, t)\right), \tag{60}$$

where ξ is a quite definite "similarity variable," $f(t)$ and $h(\xi, t)$ are known elementary functions, and G is an arbitrary function of the three indicated variables.

(ii) The requirement $Xu = 0$, i.e., that u be an absolute invariant of the generator X (incorporated in the subsidiary equations) in general implies a separation of variables in the Schrödinger equation, and we obtain a solution in the form

$$u(x, t) = R(x, t)\phi(\xi), \tag{61}$$

where $R(x, t)$ is a known elementary function and $\phi(\xi)$ is a function of the similarity variable only, satisfying an ordinary differential equation.

(iii) For all subalgebras except P, B, and E the function $\phi(\xi)$ obeys an equation of the type

$$\phi''(\xi) + W(\xi)\phi(\xi) = \phi(\xi)G(\xi, \phi, \phi^*), \tag{62}$$

where $W(\xi)$ is some specific simple function, namely one of the following types of potentials: free particle (constant), linear potential, repulsive or attractive harmonic oscillator. Thus if G depends on ξ only, we obtain a linear equation in which $W(\xi) - G(\xi)$ plays the role of a potential in a stationary Schrödinger equation. In general (62) can be called a nonlinear Schrödinger equation.

For the subalgebras P and B we obtain an equation of the type

$$i\{\phi'(t) + \alpha(t)\phi(t)\} = \phi(t)G(t, \phi, \phi^*), \tag{63}$$

where $\alpha(t) = 0$ for $X = P$ and $\alpha(t) = 1/2t$ for $X = B$.

The subalgebra $X = E$ is exceptional in that it does not provide a separation of variables. Indeed, the equation $Eu = 0$ would imply $u = 0$, and we obtain a trivial solution. The symmetry corresponding to E is, however, not trivial—it restricts the possible nonlinearity of $F = uG(x, t, |u|^2)$, i.e., G does not depend on u/u^*.

(iv) If G does not depend on u and u^*, we obtain a linear equation. The absolute invariant condition $Xu = 0$ can then be replaced by an eigenvalue type condition $Xu = ku$ with $k = $ const, and we obtain a complete set of separable solutions, rather than a single one.

Actually the linear and nonlinear cases can be treated on the same footing. The eigenvalue equation $Xu = ku$ used in conjunction with a classification of orbits of generators of the factor algebra LS_1/E is equivalent to the absolute invariant condition $Xu = 0$ used in conjunction with a classification of orbits of the entire algebra LS_1. This is completely consistent for the subalgebras $2L_3 + aE$, $2K_1 + aE$, and $2K_2 + aE$; however, for $K_2 + L_3 + \kappa E$ and $-K_2 + L_3 + \kappa E$, κ will play the role of the eigenvalue and hence we must allow it to have arbitrary values. Similarly the orbit representatives $K_2 + L_3 + \epsilon B$, $-K_2 + L_3 + \epsilon P$, P, and B must be replaced by the equivalent representatives $K_2 + L_3 + \epsilon B + aE$, $-K_2 + L_3 + \epsilon P + aE$, $P + aE$, and $B + aE$. The parameter a then plays the role of an eigenvalue.

Note that even if F is nonlinear but G has the form $G(\xi, |u|^2 f(t))$, i.e., does not depend on the phase of u, we thus obtain an infinite set of different solutions for different values of a.

(v) The separation of variables in the equation $u_{xx} + iu_t = 0$ has been studied[5] and related to orbits of generators of the factor algebra LS_1/E. Our classification of orbits is somewhat different (we classify consistently under either D or S)—and the correspondence between classes of operators and separable coordinate systems that we obtain for the equation $u_{xx} + iu_t = F$ is given in Table II. For $F = 0$ all listed systems are separable and the operator E can be omitted from the list (and we can put $a = 0$ and $\kappa = 0$).

(vi) We have included in this section all one-dimensional subalgebras listed in Table I. Each of them represents a class of subalgebras where the classification is with respect to the group D (Galilei extended by dilations). Under the group S_1 the algebras $-K_2 + L_3 + \kappa E$, $-K_2 + L_3 + \epsilon P$, and B become conjugate to $K_2 + L_3 + \kappa E$, $K_2 + L_3 + \epsilon B$, and P, respectively. Since the form of the

TABLE II. Invariant interactions and separable coordinates for one-dimensional subalgebras (the range of parameters is $-\infty < a < \infty$, $\epsilon = \pm 1$, $\kappa = 0, \pm 1$).

Diagonalized operator	Coordinates	$F(x, t, u, u^*)$
$2L_3 + aE$	$\xi = \dfrac{x}{(1+t^2)^{1/2}}$, $\eta = t$	$\dfrac{u}{1+t^2} G\left(\xi, \|u\|^2(1+t^2)^{1/2}, \dfrac{u}{u^*}\left[\dfrac{1+it}{1-it}\right]^a \exp\left[-\dfrac{i\xi^2 t}{2}\right]\right)$
$2K_1 + aE$	$\xi = \dfrac{x}{t^{1/2}}$, $\eta = t$	$\dfrac{u}{t} G\left(\xi, \|u\|^2\sqrt{t}, \dfrac{u}{u^*} t^{ia} \exp\left(-\dfrac{i\xi^2}{4}\right)\right)$
$2K_2 + aE$	$\xi = \dfrac{x}{(\|t^2-1\|)^{1/2}}$, $\eta = t$	$\dfrac{u}{\|t^2-1\|} G\left(\xi, \|u\|^2(\|t^2-1\|)^{1/2}, \dfrac{u}{u^*}\left|\dfrac{t+1}{t-1}\right|^{ia} \exp\left[-\dfrac{ix^2 t}{2(t^2-1)}\right]\right)$
$K_2 + L_3 + \epsilon B$	$\xi = x + \dfrac{\epsilon}{2}t^2$, $\eta = t$	$u\, G\left(\xi, \|u\|^2, \dfrac{u}{u^*}\exp[i(\epsilon\xi - \tfrac{1}{6}t^2)]\right)$
$-K_2 + L_3 + \epsilon P$	$\xi = \dfrac{x}{t} + \dfrac{\epsilon}{2t^2}$, $\eta = t$	$\dfrac{u}{t^2}G\left(\xi, \|u\|^2 t, \dfrac{u}{u^*}\exp\left[-\dfrac{i}{2t}\left(\xi^2 t^2 + \epsilon\xi - \dfrac{1}{12t^2}\right)\right]\right)$
$K_2 + L_3 + \kappa E$	$\xi = x$, $\eta = t$	$u\, G\left(\xi, \|u\|^2, \dfrac{u}{u^*}\exp(2i\kappa t)\right)$
$-K_2 + L_3 + \kappa E$	$\xi = \dfrac{x}{t}$, $\eta = t$	$\dfrac{u}{t^2}G\left(\xi, \|u\|^2 t, \dfrac{u}{u^*}\exp\left[-\dfrac{i}{2t}(4\kappa + \xi^2 t^2)\right]\right)$
P	$\xi = t$, $\eta = x$	$u\, G\left(t, \|u\|^2, \dfrac{u}{u^*}\right)$
B	$\xi = t$, $\eta = x$	$u\, G\left(t, \|u\|^2, \dfrac{u}{u^*}\exp\left[-\dfrac{ix^2}{2t}\right]\right)$
E	$\xi = x$, $\eta = t$	$u\, G(x, t, \|u\|^2)$

corresponding interactions, solutions, etc., are quite different, we find it worthwhile to list them separately.

(vii) The orbit representatives, separable coordinates, and invariant interactions for one-dimensional subalgebras are summarized in Table II.

C. Two-dimensional subalgebras

The results for all two-dimensional algebras listed in Table I can be obtained by combining together the results obtained for one-dimensional subalgebras. We will thus obtain less general interactions but will be able to obtain more solutions by imposing the condition $Xu = 0$ for different choices of the generator X.

Consider, for example, the algebra $\{2K_1 + aE, K_2 + L_3\}$. Invariance under $2K_1 + aE$ and $K_2 + L_3$ imply

$$F = \frac{u}{t} G\left(\frac{x}{t^{1/2}}, \|u\|^2\sqrt{t}, \frac{u}{u^*}t^{ia}\exp\left(-\frac{ix^2}{4t}\right)\right)$$

and $F = u\tilde{G}\left(x, \|u\|^2, \dfrac{u}{u^*}\right)$ (64)

respectively. These two conditions are compatible if and only if

$$F = \frac{u}{x^2}G\left(\|u\|^2 x, \frac{u}{u^*}x^{2ia}\right).$$ (65)

Nonequivalent solutions of the Schrödinger equation for the interaction (65) can be obtained by requiring $(2K_1 + aE)u = 0$ or $(K_2 + L_3)u = 0$ and will thus be of the form (42) or (44) with $\kappa = 0$.

Omitting the details, we summarize all two-dimensional algebras and the corresponding invariant interactions in Table III. Inspecting the table we see that typically F is of the form $F = (u/x^2)G$ or $F = [u/f(t)]G$, where $f(t)$

is a known function and G is an arbitrary function of two variables, of which either one or both involve u and u^*.

D. Three-dimensional subalgebras

The interactions that are invariant with respect to three-dimensional subalgebras are obtained by combining together results listed in Tables II and III. For example, consider the algebra $\{K_1, K_2, L_3\}$. Invariance under $\{K_1, K_2 + L_3\}$ implies $F = (u/x^2)G(\|u\|^2 x, u/u^*)$. Invariance under L_3 implies

$$F = \frac{u}{x^2} G\left(\frac{x}{(1+t^2)^{1/2}}, \|u\|^2 x, \frac{u}{u^*}\exp\left(-i\frac{x^2 t}{2(1+t^2)}\right)\right).$$

The intersection of these two conditions is $F = (u/x^2) \times G(\|u\|^2 x)$. Similarly we proceed with all other subalgebras. The results are summarized in Table IV. The function G will in general depend on one variable only, the only exception being the Weyl algebra $\{P, B, E\}$, leading to $F = uG(t, \|u\|^2)$.

E. Four-, five-, and six-dimensional subalgebras

The results for algebras of dimension $4 \leqslant d \leqslant 6$ are summarized in Table V.

5. CONCLUSIONS

The main results of this paper are the classification of all continuous subgroups of the Schrödinger group S_1 and the construction of an invariant interaction for each subgroup.

We should mention that the connection between subgroups and symmetry breaking interactions is not one-to-one in the considered case. Thus, for instance the interaction $F = cu\|u\|^4$ is left invariant by the entire

TABLE III. Invariant interactions for two-dimensional subalgebras (the parameters satisfy the conditions $-\infty < a < \infty$, $\kappa = 0, \pm 1$; $\epsilon = \pm 1$).

No.	Generators	Interaction
1.	$2K_1 + aE,\ K_2 + L_3,$	$F = \frac{u}{x^2} G\left(|u|^2 x, \frac{u}{u^*} x^{2ia}\right)$
2.	$2K_1 + aE,\ -K_2 + L_3,$	$F = \frac{u}{x^2} G\left(|u|^2 x, \frac{u}{u^*}\left(\frac{t}{x}\right)^{2ia} \exp\left(-\frac{ix^2}{2t}\right)\right)$
3.	$2K_1 + aE,\ P,$	$F = \frac{u}{t} G\left(|u|^2 \sqrt{t}, \frac{u}{u^*} t^{ia}\right)$
4.	$2K_1 + aE, B$	$F = \frac{u}{t} G\left(|u|^2 \sqrt{t}, \frac{u}{u^*} t^{ia} \exp\left(-\frac{ix^2}{2t}\right)\right)$
5.	$2K_2 + aE,\ B + P$	$F = \frac{u}{|t^2 - 1|} G\left(|u|^2 (|t^2 - 1|)^{1/2}, \frac{u}{u^*}\left(\frac{t+1}{t-1}\right)^{ia} \exp\left(-\frac{ix^2}{2(t-1)}\right)\right)$
6.	$2K_2 + aE,\ B - P$	$F = \frac{u}{|t^2 - 1|} G\left(|u|^2 (|t^2 - 1|)^{1/2}, \frac{u}{u^*}\left(\frac{t+1}{t-1}\right)^{ia} \exp\left(-\frac{ix^2}{2(t+1)}\right)\right)$
7.	$K_2 + L_3 + \kappa E,\ P$	$F = u\ G\left(|u|^2, \frac{u}{u^*} \exp(2i\kappa t)\right)$
8.	$-K_2 + L_3 + \kappa E,\ B$	$F = \frac{u}{t^2} G\left(|u|^2 t, \frac{u}{u^*} \exp\left[-i\left(\frac{x^2}{2t} + \frac{2\kappa}{t}\right)\right]\right)$
9.	K_1, E	$F = \frac{u}{t} G\left(\frac{x}{t^{1/2}}, |u|^2 \sqrt{t}\right)$
10.	K_2, E	$F = \frac{u}{x^2} G\left(\frac{x}{|t^2 - 1|^{1/2}}, |u|^2 |t^2 - 1|^{1/2}\right)$
11.	$K_2 + L_3,\ E$	$F = u\ G(x, |u|^2)$
12.	$K_2 + L_3 + \epsilon B,\ E$	$F = u\ G\left(x + \frac{\epsilon}{2} t^2, |u|^2\right)$
13.	$-K_2 + L_3,\ E$	$F = \frac{u}{x^2} G\left(\frac{x}{t}, |u|^2 t\right)$
14.	$-K_2 + L_3 + \epsilon P,\ E$	$F = \frac{u}{t^2} G\left(\frac{x}{t} + \frac{\epsilon}{2t^2}, |u|^2 t\right)$
15.	L_3, E	$F = \frac{u}{1 + t^2} G\left(\frac{x}{(1 + t^2)^{1/2}}, |u|^2 (1 + t^2)^{1/2}\right)$
16.	$B,\ E$	$F = u\ G(t, |u|^2)$
17.	P, E	$F = u\ G(t, |u|^2)$

group S_1, but it is also the most general type of interaction left invariant by the two five-dimensional subgroups and the two four-dimensional subgroups generated by $\{K_1, K_2 + L_3, P, E\}$ and $\{K_1, K_2 - L_3, B, E\}$. Similarly the algebras $\{K_1, K_2, L_3\}$, $\{K_1, K_2 + L_3, E\}$, $\{K_1, K_2 - L_3, E\}$, and $\{K_1 K_2 L_3 E\}$ all lead to the same type of interaction, namely $F = (u/x^2) G(|u|^2 x)$ (other such cases can be found in the tables). This is to be contrasted with the results[39] obtained for the three-dimensional stationary Schrödinger equation. For a free particle the invariance group is $E(3)$ and a one-to-one correspondence was found between subgroups of $E(3)$ and potentials of the type $V(\mathbf{r}) + A(\mathbf{r})P$ [$V(\mathbf{r})$ and $A(\mathbf{r})$ are a scalar and vector potential, P is the linear momentum operator], reducing the symmetry from $E(3)$ to the considered subgroup. Quite concievably in the time dependent case more general interactions, e.g., involving derivatives of $u(x, t)$ would be capable of distinguishing between all the subgroups.

On the positive side, let us stress that for each subgroup G_j of S_1 we have found the most general interaction $F(x, t, u, u^*)$ breaking the symmetry from S_1 to G_j.

The remaining symmetry can always be used to find a solution or many different solutions of nonlinear Schrödinger equations and a complete set of solutions for linear equations. Thus, the algebraic approach used in this paper is a method of generating solvable models.

It is of interest to notice that the equation

$$u_{xx} + iu_t = cu|u|^4$$

admits the same symmetry group as the free Schrödinger equation and thus should provide a particularly tractable model ("a ϕ^5 theory"). The term $|u|^4$ is typical for one spacelike dimension. For two spacelike dimensions the interaction would be $F = cu|u|^2$, i.e., precisely the right-hand side of the usual "nonlinear Schrödinger equation." More general powers can be obtained if the skew-hermiticity condition on the generators is dropped (as mentioned previously).

Let us make some comments on the future outlook.

(1) The one-dimensional heat equation admits a Lie group isomorphic to S_1 and hence our classification of subgroups is of relevance there. We plan to investigate

TABLE IV. Invariant interactions for three-dimensional subalgebras $(-\infty < a < \infty, \ \epsilon = \pm 1, \ \kappa = 0, \pm 1)$.

No.	Generators	Interaction
1.	K_1, K_2, L_3	$F = \frac{u}{x^2} G(|u|^2 x)$
2.	$K_1, K_2 + L_3, E$	$F = \frac{u}{x^2} G(|u|^2 x)$
3.	$K_1, K_2 - L_3, E$	$F = \frac{u}{x^2} G(|u|^2 x)$
4.	$K_1 + aE, K_2 + L_3, P$	$F = u |u|^4 G\left(\frac{u}{u^*} |u|^{-4ia}\right)$
5.	$K_1 + aE, K_2 - L_3, B$	$F = u |u|^4 G\left(\frac{u}{u^*} |u|^{4ia} t^{2ia} \exp\left(-\frac{ix^2}{2t}\right)\right)$
6.	K_1, P, E	$F = \frac{u}{t} G(|u|^2 t^{1/2})$
7.	K_1, B, E	$F = \frac{u}{t} G(|u|^2 t^{1/2})$
8.	$K_2, B+P, E$	$F = \frac{u}{|t^2-1|} G(|u|^2 |t^2-1|^{1/2})$
9.	$K_2, B-P, E$	$F = \frac{u}{|t^2-1|} G(|u|^2 |t^2-1|^{1/2})$
10.	$K_2 + L_3 + \kappa B, P, E$	$F = u G(|u|^2)$
11.	$K_2 - L_3 + \kappa P, B, E$	$F = \frac{u}{t^2} G(|u|^2 t)$
12.	P, B, E	$F = u G(t, |u|^2)$

symmetry breaking for the heat equation, where the reduction of the symmetry may be due either to additional terms in the equation, or more interestingly, due to specific boundary conditions.

(2) The results of this paper can and should be generalized to higher dimensional cases, in particular the groups S_2 and S_3 are of interest. As mentioned above, more general types of interactions can be considered, e.g., of the form $F(x, t, u, u^*, u_x, u_x^*, u_t, u_t^*, \cdots)$.

(3) We plan to make use of the existing classification of subgroups of the Poincaré group to study symmetry breaking due to external fields in the relativistic case. Again this can be considered as a source of solvable or at least tractable models for classical relativistic field theories (that may also be quantizable).

(4) A question that has not been raised, still less answered in this article, but which may be of considerable interest is the following. Given a specific linear or

TABLE V. Invariant interactions for four-, five- and six-dimensional subalgebras.

dim	No.	Generators	Interaction
4	1	K_1, K_2, L_3, E	$F = \frac{u}{x^2} G(|u|^2 x)$
	2	$K_1, K_2 + L_3, P, E$	$F = cu |u|^4$
	3	$K_1, K_2 - L_3, B, E$	$F = cu |u|^4$
	4	K_1, P, B, E	$F = \frac{u}{t} G(|u|^2 t^{1/2})$
	5	K_2, P, B, E	$F = \frac{u}{|t^2-1|} G(|t^2-1|^{1/2}|u|^2)$
	6	$K_2 + L_3, P, B, E$	$F = u G(|u|^2)$
	7	$K_2 - L_3, P, B, E$	$F = \frac{u}{t^2} G(|u|^2 t)$
	8	L_3, P, B, E	$F = \frac{u}{1+t^2} G(|u|^2 (1+t^2)^{1/2})$
5	1	$K_1, K_2 + L_3, P, B, E$	$F = cu |u|^4$
	2	$K_1, K_2 - L_3, P, B, E$	$F = cu |u|^4$
6	1	K_1, K_2, L_3, P, B, E	$F = cu |u|^4$

nonlinear Schrödinger equation, in particular one of those found in the present article, what is its complete invariance group? Generally speaking the complete invariance group may be larger than the corresponding subgroup of the Schrödinger group discussed in this paper. We plan to return to this problem in the future, specially for those nonlinear equations which promise to be of definite physical interest.

ACKNOWLEDGMENTS

The authors are much indebted to Dr. J. Patera and Dr. M. Perroud for illuminating conversations.

*Permanent address: Centro de Investigacion en Matematicas Aplicadas y en Sistemas, Universidad Nacional Autonoma de México, México 20 D. F.

†Permanent address: Department of Physics, McGill University, Montreal, Quebec, Canada.

[1]R. L. Anderson, S. Kumei, and C. E. Wulfman, Rev. Méx. Fis. 21, 1 (1972).

[2]C. P. Boyer, Helv. Phys. Acta 47, 589 (1974).

[3]U. Niederer, Helv. Phys. Acta 46, 191 (1973).

[4]G. W. Bluman and J. D. Cole, J. Math. Mech. 18, 1025 (1969).

[5]E. G. Kalnins and W. Miller, Jr., J. Math. Phys. 15, 1728 (1974).

[6]J. Patera, P. Winternitz, and H. Zassenhaus, J. Math. Phys. 15, 1378, 1932 (1974).

[7]J. Patera, P. Winternitz, and H. Zassenhaus, J. Math. Phys. 16, 1597, 1615 (1975); 17, 717 (1976).

[8]J. Patera, R. T. Sharp, P. Winternitz, and H. Zassenhaus, Can. J. Phys. 54, 950 (1976).

[9]J. Patera, R. T. Sharp, P. Winternitz, and H. Zassenhaus, J. Math. Phys. 17, 977, 986 (1976).

[10]H. Bacry, P. Combe, and P. Sorba, Rep. Math. Phys. 5, 145, 361 (1974).

[11]P. Sorba, J. Math. Phys. 17, 941 (1976).

[12]D. J. Kaup, J. Math. Phys. 16, 2036 (1975).

[13]A. C. Scott, F. Y. F. Chu, and D. McLaughlin, Proc. IEEE 61, 1443 (1973).

[14]S. Coleman, Phys. Rev. D 11, 2088 (1975).

[15]R. F. Dashen, B. Hasslacher, and A. Neveu, Phys. Rev. D 10, 4114, 4130, 4138 (1974); D 11, 3424 (1975).

[16]H. D. Wahlquist and F. B. Estabrook, J. Math. Phys. 16, 1 (1975).

[17]S. Lie, Theorie der Transformationsgruppen, Vol. 1 (1888), Vol. 2 (1890), Vol. 3 (1893) (Leipzig) (reprint by Chelsea, New York, 1967).

[18]L. V. Ovsyannikov, Gruppovye svoĭstva differentsial'nykh Uravnenii, (Novosibirsk, 1962) [Group Theoretical Properties of Differential Equations].

[19]G. Birkhoff, Hydrodynamics (Princeton U. P., Princeton, N. J., 1960).

[20]G. W. Bluman and J. D. Cole, Similarity Methods for Differential Equations (Springer, New York, 1974).

[21]W. F. Ames, Nonlinear Partial Differential Equations in Engineering Vols. 1 and 2 (Academic, New York, 1972).

[22]P. Winternitz and I. Friš, Yad. Fiz. 1, 899 (1965) [Sov. J. Nucl. Phys. 1, 636 (1965)]; P. Winternitz, I. Lukač, and Ya. A. Smorodinskiĭ, Yad. Fiz. 7, 192 (1967) [Sov. J. Nucl. Phys. 7, 139 (1968)]; A. A. Makarov, Ya. A. Smorodinskiĭ, Kh. V. Valiev, and P. Winternitz, Nuovo Cimento A 52, 1061 (1967).

[23]J. Patera and P. Winternitz, J. Math. Phys. 14, 1130 (1973).

[24]W. Miller, Jr., SIAM J. Math. Anal. 5, 626, 822 (1974).

[25]E. G. Kalnins and W. Miller, Jr., J. Math. Phys. 15, 1025, 1263 (1974).

[26]C. P. Boyer, E. G. Kalnins, and W. Miller, Jr., J. Math. Phys. 16, 499, 512 (1975).

[27]E. G. Kalnins, W. Miller, Jr., and P. Winternitz, SIAM J. Appl. Math. 30, No. 4 (1976).

[28]C. P. Boyer, SIAM J. Math. Anal. (to be published).

[29]C. P. Boyer and K. B. Wolf, J. Math. Phys. 16, 2215 (1975).

[30]W. Miller, Jr., Symmetry and Separation of Variables for Linear Differential Equations (Addison-Wesley, Reading, Mass.) (to be published).

[31]N. Ya. Vilenkin, Special Functions and the Theory of Group Representations (Amer. Math. Soc., Providence, R. I., 1968).

[32]W. Miller, Jr., Lie Theory and Special Functions (Academic, New York, 1968).

[33]J. D. Talman, Special Functions. A Group Theoretic Approach (Benjamin, New York, 1968).

[34]K. Bardakci and M. B. Halpern, Phys. Rev. 176, 1686 (1968); L. Susskind, Phys. Rev. 165, 1535 (1968); H. Bacry and N. P. Chang, Ann. Phys. (N. Y.) 47, 407 (1968); G. Burdet, M. Perrin and P. Sorba, Commun. Math. Phys. 34, 85 (1973).

[35]C. P. Boyer and M. Peñafiel, Nuovo Cimento B 31, 195 (1976).

[36]V. Bargmann, Ann. Math. 48, 569 (1947).

[37]G. M. Mubarakzyanov, Izv. Vysshikh Uchebn. Zaved. Mat. No. 1 (32) 114; No. 3 (34), 99; No. 4 (35), 104 (1963).

[38]L. P. Eisenhart, Continuous Groups of Transformations (Dover, New York, 1961).

[39]J. Beckers, J. Patera, M. Perroud, and P. Winternitz, Preprint CRM-589 (1975).

SU(2)×SU(2) scalars in the enveloping algebra of SU(4)

C. Quesne*

Physique Théorique et Mathématique, Université Libre de Bruxelles, Brussels, Belgium
(Received 26 January 1976)

We build an integrity basis for the SU(2) × SU(2) scalars belonging to the enveloping algebra of SU(4). We prove that it contains seven independent invariants in addition to the Casimir operators of SU(4) and SU(2) × SU(2). We form a complete set of commuting operators by adding to the latter two linear combinations of the former the operators Ω and Φ first introduced by Moshinsky and Nagel. We then solve the state labeling problem that occurs in the reduction SU(4) ⊃ SU(2) × SU(2) by diagonalizing simultaneously Ω and Φ. Their eigenvalues are calculated numerically in all irreducible representations of SU(4) that are encountered in light nuclei up to and including the s–d shell. Finally we build the propagation operators for the widths of the fixed supermultiplet, spin and isospin spectral distributions by taking appropriate linear combinations of SU(2) × SU(2) invariants of degree less than or equal to four, and we tabulate the averages of these operators in the above-mentioned irreducible representations of SU(4).

1. INTRODUCTION

The determination of a complete labeling for the basis states of an irreducible representation (IR) of a Lie group G, decomposed into IR's of some noncanonical subgroup H, has given rise to a lot of studies in which various types of solution have been proposed.[1-7] One of them consists in obtaining a complete set of commuting Hermitian operators by adding to the Casimir operators of G, of H, and of appropriate subgroups of H, some scalars with respect to H belonging to the enveloping algebra of G.[2,6,7] The basis states of an IR of G are then chosen as common eigenstates of this complete set, and the eigenvalues of the additional operators provide the missing labels. This type of approach of the state-labeling problem has the advantage of leading to an orthonormal basis. However, Racah has proved that it is impossible to define missing labels which have integer values for all IR's.[6]

When this procedure is adopted, it is worth while to study first the set $\mathcal{G}$ of all the scalars with respect to H, belonging to the enveloping algebra of G. In this way we can determine all possible labeling operators, and then make an appropriate choice among them.

This problem has been examined in general terms by Judd *et al.*, and a detailed application to the reduction SU(3) ⊃ O(3) has been carried out by these authors.[7] Their main result has been to show that the subalgebra $\mathcal{G}$ is finitely generated, and that an integrity basis for $\mathcal{G}$ can be built by studying first a similar problem arising in the theory of polynomial invariants. In the latter case, the construction of an integrity basis is greatly simplified by determining a generating function for the number of independent invariants of a given degree. An alternative procedure for constructing an integrity basis of $\mathcal{G}$ has been proposed recently by Sharp.[8]

Subgroup invariants in the enveloping algebra of a group also appear in another problem of physical interest, the propagation of operator averages in the spectral distribution method used in nuclear spectroscopy.[9,10] It has been shown recently[10] that when averaging over some IR's of a chain of groups, the so-called propagation operators can be written as

polynomials in the members of an integrity basis for the subgroup scalars in the enveloping algebra of the group. For the purpose of numerical calculations, it is necessary not only to write explicitly those polynomials, but also to determine their averages in all the IR's appearing in physical applications.

In the present paper we are concerned with the chain of groups SU(4) ⊃ SU(2)×SU(2) ⊃ U(1)×U(1), appearing in Wigner's supermultiplet model, in which the many-nucleon spin—isospin states are classified according to IR's of the group SU(4).[11-15] It is well known that Wigner's scheme, which rests upon a decomposition of the many-nucleon states into their space times their spin—isospin part, is a good starting basis for shell model calculations in many light nuclei up to the s—d shell.

However, the noncanonical nature of the chain of groups SU(4) ⊃ SU(2)×SU(2) gives rise to many problems when detailed applications of the model are carried out. There are two missing labels to specify the basis states of an IR of SU(4) completely, and indeed there can be more than one state characterized by given values of the spin and isospin quantum numbers S, T, M_S, and M_T in a given IR of SU(4). Moshinsky and Nagel have determined a pair of commuting, Hermitian labeling operators Ω and Φ, which solve in principle the problem.[12] However, their eigenvalues and eigenvectors are not known, and, moreover, other possible choices have not been explored. It seems therefore worth while to re-examine the problem from a general point of view.

Difficulties have also arisen when studying the fixed supermultiplet, spin, and isospin spectral distributions.[16] It has been shown that the propagation operators for the centroid energies can be constructed in terms of the Casimir operators of SU(4) and SU(2)×SU(2), but that this procedure cannot be used for the widths because other SU(2)×SU(2) invariant operators are needed.

In this paper we solve both problems by constructing an integrity basis for the SU(2)×SU(2) scalars in the enveloping algebra of SU(4). After reviewing our notations for the Lie algebras of SU(4) and SU(2)×SU(2) in Sec. 2, we derive in Sec. 3 the generating function for

the number of independent invariants of a given degree, and determine from it the explicit form of the basic scalars. We then restrict ourselves to the independent invariants of degree less than or equal to 4. The procedure used to compute their matrix elements in the Gel'fand and Tseitlin basis is described in Sec. 4. Section 5 is then devoted to the solution of the state-labeling problem. It is shown there that the pair of operators Ω and Φ, introduced by Moshinsky and Nagel, is only one among many other possible choices, but that it is quite convenient from the point of view of diagonalization. Their eigenvalues are calculated and tabulated for all the IR's of SU(4) appearing in light nuclei up to and including the $s{-}d$ shell. Finally in Sec. 6, the propagation operators for the widths of the fixed supermultiplet, spin, and isospin spectral distributions are constructed in terms of the SU(2)$\times$SU(2) invariants belonging to the previously determined integrity basis. The appropriate averages of these basic scalars are also tabulated for all the IR's of SU(4) appearing in light nuclei up to and including the $s{-}d$ shell.

2. LIE ALGEBRAS OF SU(4) AND SU(2) $\times$ SU(2)

As is well known,[17] the Lie algebra $\mathcal{U}(4)$ of the group U(4) is generated by the operators $C_\mu^{\mu'}$, $\mu, \mu' = 1, 2, 3, 4$, whose commutation relations are

$$[C_\mu^{\mu'}, C_{\mu''}^{\mu'''}] = \delta_{\mu''}^{\mu'} C_\mu^{\mu'''} - \delta_\mu^{\mu'''} C_{\mu''}^{\mu'}. \tag{2.1}$$

In terms of them, the generators of $\mathcal{U}(4)$ can be written as

$$C_\mu^{\mu'} - \tfrac{1}{4}\delta_\mu^{\mu'} \sum_{\mu''} C_{\mu''}^{\mu''}. \tag{2.2}$$

This basis of $\mathcal{U}(4)$ is adapted to the canonical chain of subgroups of SU(4), namely

$$\text{SU}(4) \supset \text{S}[\text{U}(3) \times \text{U}(1)] \supset \text{S}[\text{U}(2) \times \text{U}(1) \times \text{U}(1)]$$

$$\supset \text{S}[\text{U}(1) \times \text{U}(1) \times \text{U}(1) \times \text{U}(1)]. \tag{2.3}$$

In this paper we shall consider the noncanonical chain of subgroups

$$\text{SU}(4) \supset \text{SU}(2) \times \text{SU}(2) \supset \text{U}(1) \times \text{U}(1), \tag{2.4}$$

and use therefore another basis for the Lie algebra $\mathcal{U}(4)$, explicitly reduced with respect to the Lie algebra $\mathcal{U}(2) \times \mathcal{U}(2)$. For that purpose, we replace index μ by a double index $\sigma\tau$, where both σ and τ take two values $+\tfrac{1}{2}$ and $-\tfrac{1}{2}$, that we shall represent by $+$ and $-$ respectively, and we adopt the enumeration convention:

$$\mu \to \sigma\tau: 1 \to ++, \quad 2 \to +-, \quad 3 \to -+, \quad 4 \to --. \tag{2.5}$$

We now define the following operators[12]:

$$S_i = \tfrac{1}{2}(M_i)^\sigma_{\sigma'} (I)^\tau_{\tau'} C_{\sigma\tau}^{\sigma'\tau'},$$

$$T_\alpha = \tfrac{1}{2}(I)^\sigma_{\sigma'} (N_\alpha)^\tau_{\tau'} C_{\sigma\tau}^{\sigma'\tau'}, \quad i, \alpha = 1, 2, 3, \tag{2.6}$$

$$Q_{i\alpha} = \tfrac{1}{4}(M_i)^\sigma_{\sigma'} (N_\alpha)^\tau_{\tau'} C_{\sigma\tau}^{\sigma'\tau'},$$

where the M_i, $i = 1, 2, 3$, are the Pauli matrices associated with index σ, the N_α, $\alpha = 1, 2, 3$, those associated with index τ, and there is a summation over all dummy indices. The operators S_i, $i = 1, 2, 3$, and T_α, $\alpha = 1, 2, 3$, are the generators of $\mathcal{U}(2) \times \mathcal{U}(2)$, and, in physical terms, can be identified with the spin and isospin opera-

tors respectively. Operators (2.6) form the new basis of $\mathcal{U}(4)$ that we are looking for. Their commutation relations are

$$[S_i, S_j] = i\epsilon_{ijk}S_k, \quad [T_\alpha, T_\beta] = i\epsilon_{\alpha\beta\gamma}T_\gamma, \quad [S_i, T_\alpha] = 0,$$

$$[S_i, Q_{j\alpha}] = i\epsilon_{ijk}Q_{k\alpha}, \quad [T_\alpha, Q_{i\beta}] = i\epsilon_{\alpha\beta\gamma}Q_{i\gamma}, \tag{2.7}$$

$$[Q_{i\alpha}, Q_{j\beta}] = \tfrac{1}{4}i[\delta_{\alpha\beta}\epsilon_{ijk}S_k + \delta_{ij}\epsilon_{\alpha\beta\gamma}T_\gamma],$$

where ϵ_{ijk} and $\epsilon_{\alpha\beta\gamma}$ are the antisymmetric tensors. From (2.7), it is clear that S_i, T_α, and $Q_{i\alpha}$ transform according to the IR's $D_1 \times D_0$, $D_0 \times D_1$, and $D_1 \times D_1$ of SU(2)$\times$SU(2) respectively.

Instead of the Cartesian components of $\mathbf{S}$, $\mathbf{T}$, and $\mathbf{Q}$ given in Eqs. (2.6), it is convenient to use their spherical components

$$S_q = \tfrac{1}{2}(M_q)^\sigma_{\sigma'} (I)^\tau_{\tau'} C_{\sigma\tau}^{\sigma'\tau'},$$

$$T_\kappa = \tfrac{1}{2}(I)^\sigma_{\sigma'} (N_\kappa)^\tau_{\tau'} C_{\sigma\tau}^{\sigma'\tau'}, \quad q, \kappa = +1, 0, -1, \tag{2.8}$$

$$Q_{q\kappa} = \tfrac{1}{4}(M_q)^\sigma_{\sigma'} (N_\kappa)^\tau_{\tau'} C_{\sigma\tau}^{\sigma'\tau'},$$

where the M_q can be written as

$$M_{\pm 1} = \mp (1/\sqrt{2})[M_1 \pm iM_2],$$

$$M_0 = M_3, \tag{2.9}$$

and the N_κ are given by similar relations. The commutation relations (2.7) are now replaced by the following relations, written in terms of ordinary Wigner coefficients of SU(2):

$$[S_q, S_r] = -\sqrt{2}\langle 11qr | 1 \ q+r\rangle S_{q+r},$$

$$[T_\kappa, T_\rho] = -\sqrt{2}\langle 11\kappa\rho | 1 \ \kappa+\rho\rangle T_{\kappa+\rho},$$

$$[S_q, T_\kappa] = 0,$$

$$[S_q, Q_{r\kappa}] = -\sqrt{2}\langle 11qr | 1 \ q+r\rangle Q_{q+r\kappa}, \tag{2.10}$$

$$[T_\kappa, Q_{q\rho}] = -\sqrt{2}\langle 11\kappa\rho | 1 \ \kappa+\rho\rangle Q_{q\kappa+\rho},$$

$$[Q_{q\kappa}, Q_{r\rho}] = -\frac{1}{2\sqrt{2}} [(-1)^\kappa \delta_{\kappa,-\rho}\langle 11qr | 1 \ q+r\rangle S_{q+r}$$

$$+ (-1)^q \delta_{q,-r}\langle 11\kappa\rho | 1 \ \kappa+\rho\rangle T_{\kappa+\rho}].$$

3. SU(2) $\times$ SU(2) INVARIANTS IN THE ENVELOPING ALGEBRA OF $\mathcal{U}(4)$

We now consider the universal enveloping algebra[18] $\mathcal{A}$ of $\mathcal{U}(4)$, and we proceed to determine the set $\mathcal{J}$ of elements of $\mathcal{A}$ which are left invariant under the action of SU(2)$\times$SU(2).

When considered as a vector space, the associative algebra $\mathcal{J}$ can be written in the form

$$\mathcal{J} = \sum_{m=0}^{\infty} \oplus \mathcal{J}_m, \tag{3.1}$$

where $\mathcal{J}_m$ is the space of all symmetric polynomials $p_m(X_1, \ldots, X_{15})$ in the $\mathcal{U}(4)$ generators X_i, $i = 1, \ldots, 15$ [given for instance in Eqs. (2.6) or (2.8)], which are homogeneous of degree m and are left invariant under the action of SU(2)$\times$SU(2).

Judd et al.[7] have given general rules for the construction of $\mathcal{J}$ in the case where a connected Lie group G is reduced to a Lie subgroup H. They can be applied here if we make $G = \text{SU}(4)$, and $H = \text{SU}(2) \times \text{SU}(2)$. The main point of their study has been to show that it is

sufficient to build an integrity basis of $\mathcal{G}$, and that this construction is always possible. Following their definition, an integrity basis of $\mathcal{G}$ is a minimal generating subset of $\mathcal{G}$, or in other words a finite set of invariants $\{i_1, \ldots, i_q\}$ such that: (1) Each $i_j \in \mathcal{G}$ is homogeneous of degree $m_j \geq 1$ and symmetric in $X_1, \ldots, X_{15}$, i.e., each $i_j \in \mathcal{G}_{m_j}$. (2) Every $i \in \mathcal{G}$ can be expressed as a polynomial in $i_1, \ldots, i_q$. (3) No one of the i_k can be expressed as a polynomial in the remaining i_j, $j \neq k$. Let us note that the condition of symmetry of each i_j can be relaxed, this giving rise to an integrity basis more suitable to numerical calculations.

The construction of an integrity basis of $\mathcal{G}$ is made possible by establishing a link with a similar problem in the theory of polynomial invariants. Let us consider column vectors $\mathbf{x} = (x_1, \ldots, x_{15})$, which form a basis for the 15×15 matrix representation of SU(2)×SU(2), which is induced by the adjoint action of SU(2)×SU(2) on the basis $X_1, \ldots, X_{15}$ of $SU(4)$, and let $I[\mathbf{x}]$ be the associative algebra whose elements are the polynomials in the indeterminants $x_1, \ldots, x_{15}$ over the complex field, which are left invariant under the action of SU(2)×SU(2). It is well known[19] that $I[\mathbf{x}]$ has a finite integrity basis whose definition is similar to that of $\mathcal{G}$. Judd et al.[7] have shown that if $\{i_1(\mathbf{x}), \ldots, i_r(\mathbf{x})\}$ is an integrity basis of $I[\mathbf{x}]$, then the set $\{i_1(X_1, \ldots, X_{15}), \ldots, i_r(X_1, \ldots, X_{15})\}$ contains an integrity basis of $\mathcal{G}$. In general the set $\{i_1(X_1, \ldots, X_{15}), \ldots, i_r(X_1, \ldots, X_{15})\}$ is not itself an integrity basis of $\mathcal{G}$ because the commutation relations of $X_1, \ldots, X_{15}$ are responsible for some algebraic relations between $i_1(X_1, \ldots, X_{15}), \ldots, i_r(X_1, \ldots, X_{15})$ in $\mathcal{G}$, which have no counterpart in $I[\mathbf{x}]$. Thus, to find an integrity basis of $\mathcal{G}$, it is sufficient to find first an integrity basis $\{i_1, \ldots, i_r\}$ of $I[\mathbf{x}]$, and then to form all possible commutators $[i_l(X_j), i_m(X_j)]$ in order to determine a minimal subset of the $i_k(X_j)$ which are independent. We now proceed to implement this program.

Let us consider the space of all polynomials in the 15 indeterminants $s_i, t_\alpha, q_{i\alpha}$, $i, \alpha = 1, 2, 3$, which transform under SU(2)×SU(2) according to the IR's $D_1 \times D_0$, $D_0 \times D_1$, and $D_1 \times D_1$ respectively. The subspace $P_{A, B, C}[s_i, t_\alpha, q_{i\alpha}]$ of homogeneous polynomials of degree A, B, and C in the s_i, t_α, and $q_{i\alpha}$ respectively, is clearly invariant under the action of SU(2)×SU(2). Thus we can classify polynomial invariants $C^{(A\,B\,C)}$ in terms of their degrees of homogeneity A, B, C in the s_i, t_α, and $q_{i\alpha}$.

Following a technique developed in Ref. 7, it is possible to determine the number $N_{A, B, C}$ of invariants of degrees A, B, C in $P[s_i, t_\alpha, q_{i\alpha}]$ by deriving a generating function for it. For that purpose, we need the characters $\chi_{S, T}(\theta_1, \theta_2)$ of the IR's $D_S \times D_T$ of SU(2)×SU(2), which are given by[20]

$$\chi_{S, T}(\theta_1, \theta_2) = \sum_{M_S=-S}^{S} \sum_{M_T=-T}^{T} \exp[i(M_S\theta_1 + M_T\theta_2)], \quad (3.2)$$

and satisfy the well-known orthogonality relations

$$\pi^{-2} \int_0^{2\pi} d\theta_1 \sin^2\frac{\theta_1}{2} \int_0^{2\pi} d\theta_2 \sin^2\frac{\theta_2}{2} \chi_{S', T'}^*(\theta_1, \theta_2)\chi_{S, T}(\theta_1, \theta_2)$$
$$= \delta_{SS'}\delta_{TT'}. \quad (3.3)$$

Let us denote now by $\chi_{A, B, C}(\theta_1, \theta_2)$ the (compound) character of the representation of SU(2)×SU(2) whose

representation space is the subspace $P_{A, B, C}[s_i, t_\alpha, q_{i\alpha}]$. It is clearly given by the relation

$$\chi_{A, B, C}(\theta_1, \theta_2) = \sum_{a, \ldots, p} \exp[i\theta_1(a - c + g + h + i - m - n - p)$$
$$+ i\theta_2(d - f + g - i + j - l + m - p)], \quad (3.4)$$

where the sum is taken over all nonnegative integers $a, \ldots, p$ such that $a + b + c = A$, $d + e + f = B$, and $g + h + i + j + k + l + m + n + p = C$. From relation (3.4), it follows that

$$F[\exp(i\theta_1), \exp(i\theta_2), x, y, z]$$

$$\equiv \{[1 - \exp(i\theta_1)x](1 - x)[1 - \exp(-i\theta_1)x][1 - \exp(i\theta_2)y]$$

$$\times (1 - y)[1 - \exp(-i\theta_2)y][1 - \exp(i\theta_1 + i\theta_2)z]$$

$$\times [1 - \exp(i\theta_1)z][1 - \exp(i\theta_1 - i\theta_2)z][1 - \exp(i\theta_2)z]$$

$$\times (1 - z)[1 - \exp(-i\theta_2)z][1 - \exp(-i\theta_1 + i\theta_2)z]$$

$$\times [1 - \exp(-i\theta_1)z][1 - \exp(-i\theta_1 - i\theta_2)z]\}^{-1} \quad (3.5)$$

is a generating function for the character $\chi_{A, B, C}(\theta_1, \theta_2)$, i.e.,

$$F[\exp(i\theta_1), \exp(i\theta_2), x, y, z]$$

$$= \sum_{A, B, C=0}^{\infty} \chi_{A, B, C}(\theta_1, \theta_2)x^A y^B z^C. \quad (3.6)$$

As $N_{A, B, C}$ is the multiplicity of the identity representation $D_0 \times D_0$ of SU(2)×SU(2) in $P_{A, B, C}[s_i, t_\alpha, q_{i\alpha}]$, we obtain from Eqs. (3.3) and (3.6) that

$$\pi^{-2} \int_0^{2\pi} d\theta_1 \sin^2\frac{\theta_1}{2} \int_0^{2\pi} d\theta_2 \sin^2\frac{\theta_2}{2}$$

$$\times F[\exp(i\theta_1), \exp(i\theta_2), x, y, z]$$

$$= \sum_{A, B, C=0}^{\infty} N_{A, B, C}x^A y^B z^C \quad (3.7)$$

is a generating function for the number of invariants $N_{A, B, C}$. It remains now to perform both integrations contained in Eq. (3.7). For that purpose, we set $\lambda_1 = \exp(i\theta_1)$ and $\lambda_2 = \exp(i\theta_2)$, and convert the left-hand side of Eq. (3.7) into two successive contour integrals about a unit circle in the complex plane, that we can evaluate by the calculus of residues. After straightforward, but lengthy calculations, we get the following result[21]:

$$\sum_{A, B, C=0}^{\infty} N_{A, B, C}x^A y^B z^C$$

$$= [(1 - x^2)(1 - y^2)(1 - z^2)(1 - xyz)(1 - z^3)$$

$$\times (1 - x^2z^2)(1 - y^2z^2)(1 - z^4)(1 - x^2z^4)(1 - y^2z^4)]^{-1}$$

$$\times \{1 + xyz^2 + xyz^3 + (x^2y + xy^2)z^3 + (x^2y + xy^2)z^4$$

$$+ (x^2y + xy^2)z^5 + [(x^3 + y^3)z^6 + x^2y^2z^5]$$

$$- [x^2y^2z^8 + (x^4y + xy^4)z^7] - (x^3y^2 + x^2y^3)z^8$$

$$- (x^3y^2 + x^2y^3)z^9 - (x^3y^2 + x^2y^3)z^{10} - x^3y^3z^{10}$$

$$- x^3y^3z^{11} - x^4y^4z^{13}\}. \quad (3.8)$$

The denominator of Eq. (3.8) is produced by all the polynomials in the independent invariants

$$C^{(200)} = s_i s_i, \qquad\qquad C^{(020)} = t_\alpha t_\alpha,$$

$$C^{(002)} = q_{i\alpha}q_{i\alpha}, \qquad\qquad C^{(111)} = s_i t_\alpha q_{i\alpha},$$

$$C^{(003)} = \epsilon_{ijk}\epsilon_{\alpha\beta\gamma}q_{i\alpha}q_{j\beta}q_{k\gamma}, \quad C^{(202)} = s_i s_j q_{i\alpha}q_{j\alpha},$$

$$C^{(022)} = t_\alpha t_\beta q_{i\alpha}q_{i\beta}, \qquad C^{(004)} = q_{i\alpha}q_{i\beta}q_{j\alpha}q_{j\beta}, \qquad (3.9)$$

$$C^{(204)} = s_i s_j q_{i\alpha}q_{j\beta}q_{k\alpha}q_{k\beta}, \quad C^{(024)} = t_\alpha t_\beta q_{i\alpha}q_{j\beta}q_{i\gamma}q_{j\gamma},$$

which belong therefore to an integrity basis of $I[s_i, t_\alpha, q_{i\alpha}]$.

To explain the positive terms in the numerator of Eq. (3.8), it is necessary to add to them the invariants

$$C^{(112)} = \epsilon_{ijk}\epsilon_{\alpha\beta\gamma}s_i t_\alpha q_{j\beta}q_{k\gamma},$$

$$C^{(113)} = s_i t_\alpha q_{i\beta}q_{j\alpha}q_{j\beta},$$

$$C^{(213)} = \epsilon_{ijk}s_i s_l t_\alpha q_{j\alpha}q_{k\beta}q_{l\beta},$$

$$C^{(123)} = \epsilon_{\alpha\beta\gamma}s_i t_\alpha t_\delta q_{i\beta}q_{j\gamma}q_{j\delta},$$

$$C^{(214)} = \epsilon_{\alpha\beta\gamma}s_i s_j t_\alpha q_{i\beta}q_{j\delta}q_{k\gamma}q_{k\delta},$$

$$C^{(124)} = \epsilon_{ijk}s_i t_\alpha t_\beta q_{j\alpha}q_{l\beta}q_{k\gamma}q_{l\gamma}, \qquad (3.10)$$

$$C^{(215)} = \epsilon_{ijk}s_i s_l t_\alpha q_{j\alpha}q_{k\beta}q_{l\gamma}q_{m\beta}q_{m\gamma},$$

$$C^{(125)} = \epsilon_{\alpha\beta\gamma}s_i t_\alpha t_\delta q_{i\beta}q_{j\gamma}q_{k\delta}q_{k\epsilon}q_{j\epsilon},$$

$$C^{(306)} = \epsilon_{ijk}s_i s_l s_m q_{j\alpha}q_{k\beta}q_{l\alpha}q_{m\gamma}q_{n\beta}q_{n\gamma},$$

and

$$C^{(036)} = \epsilon_{\alpha\beta\gamma}t_\alpha t_\delta t_\epsilon q_{i\beta}q_{j\gamma}q_{i\delta}q_{k\epsilon}q_{j\varphi}q_{k\varphi}.$$

It is easy to check that $C^{(113)}$, $C^{(215)}$, and $C^{(125)}$, as defined in Eqs. (3.10), are independent of $C^{(111)}C^{(002)}$, $C^{(213)}C^{(002)}$, and $C^{(123)}C^{(002)}$, respectively, as they should be. On the other hand, the product $C^{(113)}C^{(112)}$ is independent of $C^{(202)}C^{(003)}C^{(020)}$, $C^{(022)}C^{(003)}C^{(200)}$, $C^{(112)}C^{(111)}C^{(002)}$, $[C^{(111)}]^2 C^{(003)}$, and $C^{(003)}C^{(200)}C^{(020)}$ $\times C^{(002)}$, and thus gives rise to the term $x^2 y^2 z^5$ of the generating function, without necessity of introducing a new invariant $C^{(225)}$.

All the powers of the invariants given in Eqs. (3.10) do not define independent invariants. The same is true for the products of powers of the invariants given in Eqs. (3.9) with those given in Eqs. (3.10). For instance $[C^{(112)}]^2$ can be expressed as a polynomial in the invariants (3.9) and $C^{(113)}$, which is of first degree in $C^{(113)}$:

$$[C^{(112)}]^2 = 4[C^{(204)}C^{(020)} + C^{(024)}C^{(200)}] - 8C^{(113)}C^{(111)}$$

$$+ 4C^{(202)}C^{(022)} - 4[C^{(202)}C^{(020)} + C^{(022)}C^{(200)}]C^{(002)}$$

$$- 2C^{(004)}C^{(200)}C^{(020)} + 4[C^{(111)}]^2 C^{(002)}$$

$$+ 2C^{(200)}C^{(020)}[C^{(002)}]^2. \qquad (3.11)$$

The existence of such a relation is exhibited by the absence of a term $x^2 y^2 z^4$ in the numerator of the generating function. The other relations between powers of invariants involve polynomials of degree greater than or equal to 10, and are responsible for the negative terms in the numerator of the generating function. Owing to their high degree of complexity, we have not explored them further.

When replacing s_i, t_α, and $q_{i\alpha}$ by S_i, T_α, and $Q_{i\alpha}$ in the integrity basis of $I[s_i, t_\alpha, q_{i\alpha}]$ defined in Eqs. (3.9) and (3.10), we get a set of 20 SU(2)×SU(2) invariants belonging to the enveloping algebra of $\mathcal{SU}(4)$. The Casimir operators G_2, G_3, G_4 of SU(4), and $\mathbf{S}^2, \mathbf{T}^2$ of SU(2)×SU(2) are, of course, SU(2)×SU(2) invariants

and can thus be written in terms of them. We find indeed that

$$\mathbf{S}^2 = C^{(200)}, \quad \mathbf{T}^2 = C^{(020)},$$

$$G_2 = C^{(200)} + C^{(020)} + 4C^{(002)},$$

$$G_3 = 2G_2 + 6C^{(111)} - 4C^{(003)},$$

and $\qquad\qquad\qquad\qquad\qquad\qquad (3.12)$

$$G_4 = 4G_3 - 3G_2 + 4C^{(202)} + 4C^{(022)} - 4C^{(112)} - 8C^{(004)}$$

$$+ \tfrac{1}{4}[C^{(200)}]^2 + \tfrac{3}{2}C^{(200)}C^{(020)} + 2C^{(200)}C^{(002)}$$

$$+ \tfrac{1}{4}[C^{(020)}]^2 + 2C^{(020)}C^{(002)} + 12[C^{(002)}]^2$$

$$- C^{(200)} + \tfrac{1}{2}C^{(020)}.$$

We include therefore the Casimir operators in the above-mentioned set of invariants by dropping $C^{(200)}$, $C^{(020)}$, $C^{(002)}$, $C^{(003)}$, and $C^{(004)}$.

It remains now to determine which invariants remain algebraicly independent when the commutation relations (2.7) are taken into account. It is easy to check that the operators $C^{(213)}$, $C^{(123)}$, $C^{(214)}$, $C^{(124)}$, $C^{(215)}$, $C^{(125)}$, $C^{(306)}$, and $C^{(036)}$ can be expressed in terms of the commutators $[C^{(111)}, C^{(202)}]$, $[C^{(111)}, C^{(022)}]$, $[C^{(202)}, C^{(112)}]$, $[C^{(022)}, C^{(112)}]$, $[C^{(202)}, C^{(113)}]$, $[C^{(022)}, C^{(113)}]$, $[C^{(202)}, C^{(204)}]$, and $[C^{(022)}, C^{(024)}]$, respectively, and lower order terms. In addition to the Casimir operators G_2, G_3, G_4, $\mathbf{S}^2$, and $\mathbf{T}^2$, the integrity basis for the SU(2)×SU(2) invariants in the enveloping algebra of $\mathcal{SU}(4)$, that we have built here, contains therefore seven independent invariants, $C^{(111)}$, $C^{(202)}$, $C^{(022)}$, $C^{(112)}$, $C^{(113)}$, $C^{(204)}$, and $C^{(024)}$.

In order to be able to apply this result to physical problems, we need a method to calculate the matrix elements of the basic invariants in all IR's of SU(4) appearing in applications. As the matrix elements of the Casimir operators are well known, we are left with the determination of those of the other seven invariants. In the next section, we study the properties of the matrix elements of the four invariants of degree less than or equal to 4, and show how they can be calculated.

4. MATRIX ELEMENTS OF THE SU(2) × SU(2) INVARIANTS OF DEGREE LESS THAN OR EQUAL TO 4 IN THE CANONICAL CHAIN OF SUBGROUPS OF SU(4)

To calculate the matrix elements of $C^{(111)}$, $C^{(202)}$, $C^{(022)}$, and $C^{(112)}$, it is convenient to replace the Cartesian components of S, T, and Q by their spherical ones, defined in Eqs. (2.8). The invariant operators become

$$C^{(111)} = \sum_{q\kappa} (-1)^{q+\kappa}S_q T_\kappa Q_{-q-\kappa},$$

$$C^{(202)} = \sum_{qq'\kappa} (-1)^{q+q'+\kappa}S_q S_{q'}Q_{-q\kappa}Q_{-q'-\kappa},$$

$$C^{(022)} = \sum_{q\kappa\kappa'} (-1)^{q+\kappa+\kappa'}T_\kappa T_{\kappa'}Q_{q-\kappa}Q_{-q-\kappa'}, \qquad (4.1)$$

$$C^{(112)} = -\sum_{\substack{qq'q'' \\ \kappa\kappa'\kappa''}} \epsilon_{qq'q''}\epsilon_{\kappa\kappa'\kappa''}S_q T_\kappa Q_{q'\kappa'}Q_{q''\kappa''}.$$

After expanding the sums in Eqs. (4.1) and using Eqs. (2.10) to permute the generators of SU(4), we get

$$C^{(111)} = [S_{-1}T_{-1}Q_{11} - S_{-1}T_0Q_{10} + S_{-1}T_1Q_{1-1} - (S_0+1)T_{-1}Q_{01}$$
$$+ \text{h. c.}] + (S_0+1)(T_0+1)Q_{00},$$

$$C^{(202)} = [S_{-1}^2(-2Q_{1-1}Q_{11} + Q_{10}^2) + 2(S_0+1)S_{-1}(Q_{01}Q_{1-1}$$
$$- Q_{10}Q_{00} + Q_{0-1}Q_{11}) + S_{-1}S_1(-Q_{-1-1}Q_{11} + Q_{-10}Q_{10}$$
$$- Q_{-11}Q_{1-1} - \tfrac{1}{8}S_0 + \tfrac{5}{8}) + \text{h. c.}] + (S_0+1)^2$$
$$\times (-2Q_{0-1}Q_{01} + Q_{00}^2 + \tfrac{1}{4}T_0) + (S_0+1)$$
$$\times (-Q_{-1-1}Q_{11} + Q_{-10}Q_{10} - Q_{-11}Q_{1-1} - \tfrac{3}{4}S_0),$$

$$C^{(022)} = [T_{-1}^2(-2Q_{-11}Q_{11} + Q_{01}^2) + 2(T_0+1)T_{-1} \qquad (4.2)$$
$$\times (Q_{-11}Q_{10} - Q_{01}Q_{00} + Q_{-10}Q_{11}) + T_{-1}T_1$$
$$\times (-Q_{-1-1}Q_{11} + Q_{0-1}Q_{01} - Q_{-11}Q_{1-1} - \tfrac{1}{4}S_0$$
$$- \tfrac{3}{8}T_0 + \tfrac{1}{8}) + \text{h. c.}] + (T_0+1)^2(-2Q_{-10}Q_{10} + Q_{00}^2)$$
$$+ (T_0+1)[-Q_{-1-1}Q_{11} + Q_{0-1}Q_{01} - Q_{-11}Q_{1-1}$$
$$+ \tfrac{1}{4}(S_0-2)T_0],$$

and

$$C^{(112)} = -2[S_{-1}T_{-1}(Q_{11}Q_{00} - Q_{01}Q_{10}) + S_{-1}T_0(Q_{01}Q_{1-1} - Q_{0-1}Q_{11})$$
$$+ S_{-1}T_1(Q_{0-1}Q_{10} - Q_{1-1}Q_{00}) + (S_0+1)T_{-1}(Q_{-11}Q_{10}$$
$$- Q_{-10}Q_{11}) + \text{h. c.}] + (S_0+1)(T_0+1)(Q_{-1-1}Q_{11}$$
$$- Q_{-11}Q_{1-1}) + \tfrac{1}{2}(S_0+1)(2T_{-1}T_1 - T_0^2 - T_0),$$

where h. c. stands for the Hermitian conjugate of the preceding terms.

We want now to evaluate the matrix elements of these operators between basis states of an IR $[\mathbf{m}] = [m_1 m_2 m_3]$ of SU(4). Here m_1, m_2, and m_3 are any integers, satisfying the inequalities $m_1 \geq m_2 \geq m_3 \geq 0$. We shall adopt the so-called Gel'fand and Tseitlin basis,[23] corresponding to the canonical chain of subgroups of SU(4), given in Eq. (2.3). The basis states are then represented by the patterns

$$\left| \begin{matrix} m_{14} & m_{24} & m_{34} & m_{44} \\ & m_{13} & m_{23} & m_{33} \\ & & m_{12} & m_{22} \\ & & & m_{11} \end{matrix} \right\rangle, \qquad (4.3)$$

where $m_{i4} = m_i$, $i = 1, 2, 3$, $m_{44} = 0$, and m_{ik} are integers such that $m_{i, k+1} \geq m_{ik} \geq m_{i+1, k}$.

The basis states (4.3) are simultaneous eigenvectors of a complete set of commuting operators, consisting of the Casimir operators G_2, G_3, and G_4 of SU(4) and all the Casimir operators of the subgroups appearing in Eq. (2.3). In particular they are simultaneous eigenvectors of C_{++}^{++}, C_{+-}^{+-}, C_{-+}^{-+}, and C_{--}^{--}, corresponding to the eigenvalues m_{11}, $m_{12} + m_{22} - m_{11}$, $m_{13} + m_{23} + m_{33}$ $- (m_{12} + m_{22})$, and $m_{14} + m_{24} + m_{34} + m_{44} - (m_{13} + m_{23}$ $+ m_{33})$, respectively. Since S_0, T_0, and Q_{00} are linear combinations of them, they are also diagonal, and their eigenvalues are equal to

$$M_S = m_{12} + m_{22} - \tfrac{1}{2}(m_{14} + m_{24} + m_{34} + m_{44}), \qquad (4.4a)$$

$$M_T = m_{11} - (m_{12} + m_{22}) + m_{13} + m_{23} + m_{33}$$
$$- \tfrac{1}{2}(m_{14} + m_{24} + m_{34} + m_{44}), \qquad (4.4b)$$

and

$$M_Q = \tfrac{1}{2}[m_{11} - (m_{13} + m_{23} + m_{33}) + \tfrac{1}{2}(m_{14} + m_{24} + m_{34} + m_{44})], \qquad (4.4c)$$

respectively. However, the Casimir operators of $SU(2) \times SU(2)$, $\mathbf{S}^2$ and $\mathbf{T}^2$, are not diagonal in basis (4.3).

The matrix elements of $C^{(111)}$, $C^{(202)}$, $C^{(022)}$, and $C^{(112)}$ are easily obtained in basis (4.3) when use is made of Eqs. (4.2), the expansions (2.8) of the operators S_q, T_κ, $Q_{q\kappa}$ in terms of $C_{\sigma\tau}^{\sigma'\tau'}$, and the well-known matrix elements of $C_{\sigma\tau}^{\sigma'\tau'}$ in the Gel'fand and Tseitlin basis.[23,24] Proceeding in this way, we are able to compute them numerically for all SU(4) IR's appearing in physical applications.

The calculations can be simplified by taking into account a symmetry property of the matrix elements. It is indeed straightforward to show that the matrix elements of the generators of SU(4) in a given IR are related to those in the contragredient IR as follows:

$$\left\langle \begin{matrix} m_{14}-m_{44} & m_{14}-m_{34} & m_{14}-m_{24} & 0 \\ m_{14}-m_{33}' & m_{14}-m_{23}' & m_{14}-m_{13}' & \\ m_{14}-m_{22}' & m_{14}-m_{12}' & & \\ m_{14}-m_{11}' & & & \end{matrix} \left| \begin{Bmatrix} S_q \\ T_\kappa \\ Q_{q\kappa} \end{Bmatrix} \right| \begin{matrix} m_{14}-m_{44} & m_{14}-m_{34} & m_{14}-m_{24} & 0 \\ m_{14}-m_{33} & m_{14}-m_{23} & m_{14}-m_{13} & \\ m_{14}-m_{22} & m_{14}-m_{12} & & \\ m_{14}-m_{11} & & & \end{matrix} \right\rangle$$

$$= \varphi \left\langle \begin{matrix} m_{14} & m_{24} & m_{34} & m_{44} \\ m_{13}' & m_{23}' & m_{33}' & \\ m_{12}' & m_{22}' & & \\ m_{11}' & & & \end{matrix} \left| \begin{Bmatrix} S_{-q} \\ T_{-\kappa} \\ Q_{-q-\kappa} \end{Bmatrix} \right| \begin{matrix} m_{14} & m_{24} & m_{34} & m_{44} \\ m_{13} & m_{23} & m_{33} & \\ m_{12} & m_{22} & & \\ m_{11} & & & \end{matrix} \right\rangle, \qquad (4.5)$$

where φ is a phase equal to $+1$ for $S_{\pm 1}$, $Q_{\pm 1\kappa}$, and -1 for S_0, T_κ, $Q_{0\kappa}$. From this we deduce that

$$\left\langle \begin{matrix} m_{14}-m_{44} & m_{14}-m_{34} & m_{14}-m_{24} & 0 \\ m_{14}-m_{33}' & m_{14}-m_{23}' & m_{14}-m_{13}' & \\ m_{14}-m_{22}' & m_{14}-m_{12}' & & \\ m_{14}-m_{11}' & & & \end{matrix} \left| \begin{Bmatrix} C^{(111)} \\ C^{(202)} \\ C^{(022)} \\ C^{(112)} \end{Bmatrix} \right| \begin{matrix} m_{14}-m_{44} & m_{14}-m_{34} & m_{14}-m_{24} & 0 \\ m_{14}-m_{33} & m_{14}-m_{23} & m_{14}-m_{13} & \\ m_{14}-m_{22} & m_{14}-m_{12} & & \\ m_{14}-m_{11} & & & \end{matrix} \right\rangle$$

$$= \begin{Bmatrix} -1 \\ +1 \\ +1 \\ +1 \end{Bmatrix} \left\langle \begin{matrix} m_{14} & m_{24} & m_{34} & m_{44} \\ m_{13}' & m_{23}' & m_{33}' & \\ m_{12}' & m_{22}' & & \\ m_{11}' & & & \end{matrix} \left| \begin{Bmatrix} C^{(111)} \\ C^{(202)} \\ C^{(022)} \\ C^{(112)} \end{Bmatrix} \right| \begin{matrix} m_{14} & m_{24} & m_{34} & m_{44} \\ m_{13} & m_{23} & m_{33} & \\ m_{12} & m_{22} & & \\ m_{11} & & & \end{matrix} \right\rangle. \qquad (4.6)$$

Finally let us note that the matrix elements of the invariants in any U(4) IR $[\mathbf{m}] = [m_1 m_2 m_3 m_4]$ coincide with those in the SU(4) IR $[m_1 - m_4 \; m_2 - m_4 \; m_3 - m_4]$.

We proceed now to review two applications of the invariant operators in mathematical and physical problems.

5. APPLICATION TO THE STATE-LABELING PROBLEM

Instead of using the canonical chain of groups (2.3) considered in Sec. 4, we are in fact concerned with the noncanonical chain (2.4), which is of greater physical interest. Basis states of SU(4), corresponding to the latter chain, are simultaneous eigenvectors of G_2, G_3, G_4, S^2, T^2, S_0, and T_0. There are two missing labels to characterize the states completely. Consequently, the number $d([m_1 m_2 m_3] \; ST)$ of IR's of SU(2)×SU(2), specified by S and T, in a given IR of SU(4), characterized by $[m_1 m_2 m_3]$, may be greater than one. This fact is illustrated in Tables I and II, which give the decomposition of all the IR's of SU(4) appearing in $s-d$ shell nuclei. The tables were constructed from the corresponding tables[25] for the reduction

$$U(4) \supset O(4),$$
$$\;\; [\mathbf{m}] \qquad [\lambda\mu]$$
$$(5.1)$$

using the following properties: (i) The IR $[\lambda\mu]$ of O(4) is an IR with respect to SO(4) when $\mu = 0$, and separates into two IR's of SO(4), characterized by $[\lambda\mu]$ and $[\lambda - \mu]$ respectively, when $\mu \neq 0$. (ii) The IR $[\lambda\mu]$ of

SO(4) is an IR of the locally isomorphic SU(2)×SU(2) group, characterized by $S = \frac{1}{2}(\lambda + \mu)$, and $T = \frac{1}{2}(\lambda - \mu)$.

The two missing label operators should be commuting SU(2)×SU(2) scalars in the enveloping algebra of $\mathcal{SU}(4)$. Moshinsky and Nagel have shown that they can be chosen as

$$\Omega \equiv C^{(111)},$$
$$(5.2)$$

and

$$\Phi \equiv C^{(202)} + C^{(022)} - C^{(112)}.$$
$$(5.3)$$

From Sec. 3, it is clear that this choice is only one among numerous other possibilities. The most convenient one, from the point of view of diagonalization, corresponds obviously to invariants of lowest degree. If we restrict to homogeneous polynomial invariants, we could thus take any linear combinations of the type $\alpha_1 G_3 + \alpha_2 C^{(111)}$, and $\beta_1 G_4 + \beta_2 G_2^2 + \beta_3 S^4 + \beta_4 T^4 + (\beta_5 S^2 + \beta_6 T^2)G_2 + \beta_7 S^2 T^2 + \beta_8(C^{(202)} + C^{(022)} - C^{(112)})$. The choice made by Moshinsky and Nagel belongs to this class, and is therefore well adapted to the numerical calculation of the missing labels, that we now proceed to discuss.

TABLE I. Decomposition of the IR's $[m]$ of SU(4), with $\sum_i m_i$ even, into IR's of SU(2)×SU(2), characterized by S and T.

$[m]$	$(2S, 2T)^{d([m]ST)}$
[0]	(00)
[2]	(00) (22)
[11]	(20) (02)
[4]	(00) (22) (44)
[31]	(20) (02) (22) (42) (24)
[22]	(00) (40) (22) (04)
[211]	(20) (02) (22)
[6]	(00) (22) (44) (66)
[51]	(20) (02) (22) (42) (24) (44) (64) (46)
[42]	(00) (40) (22)^2 (42) (62) (04) (24) (44) (26)
[411]	(20) (02) (22) (42) (24) (44)
[33]	(20) (60) (02) (42) (24) (06)
[321]	(20) (40) (02) (22)^2 (42) (04) (24)
[62]	(00) (40) (22)^2 (42) (62) (04) (24) (44)^2 (64) (84) (26) (46) (66) (48)
[611]	(20) (02) (22) (42) (24) (44) (64) (46) (66)
[53]	(20) (60) (02) (22) (42)^2 (62) (82) (24)^2 (44) (64) (06) (26) (46) (28)
[521]	(20) (40) (02) (22)^2 (42)^2 (62) (04) (24)^2 (44)^2 (64) (26) (46)
[44]	(00) (40) (80) (22) (62) (04) (44) (26) (08)
[431]	(20) (40) (60) (02) (22)^2 (42)^2 (62) (04) (24)^2 (44) (06) (26)
[422]	(00) (40) (22)^2 (42) (04) (24) (44)
[64]	(00) (40) (80) (22)^2 (42) (62)^2 (82) (10,2) (04) (24) (44)^2 (64) (84) (26)^2 (46) (66) (08) (28) (48) (2,10)
[631]	(20) (40) (60) (02) (22)^2 (42)^3 (62)^2 (82) (04) (24)^3 (44)^3 (64)^2 (84) (06) (26)^2 (46)^2 (66) (28) (48)
[622]	(00) (40) (22)^2 (42) (62) (04) (24) (44)^2 (64) (26) (46) (66)
[55]	(20) (60) (10,0) (02) (42) (82) (24) (64) (06) (46) (28) (0,10)
[541]	(20) (40) (60) (80) (02) (22)^2 (42)^2 (62)^2 (82) (04) (24)^2 (44)^2 (64) (06) (26)^2 (46) (08) (28)
[532]	(20) (40) (60) (02) (22)^2 (42)^3 (62) (04) (24)^3 (44)^2 (64) (06) (26) (46)
[66]	(00) (40) (80) (12,0) (22) (62) (10,2) (04) (44) (84) (26) (66) (08) (48) (2,10) (0,12)
[651]	(20) (40) (60) (80) (10,0) (02) (22)^2 (42)^2 (62)^2 (82)^2 (10,2) (04) (24)^2 (44)^2 (64)^2 (84) (06) (26)^2 (46)^2 (66) (08) (28)^2 (48) (0,10) (2,10)
[642]	(00) (40)^2 (60) (80) (22)^3 (42)^3 (62)^3 (82) (04)^2 (24)^3 (44)^4 (64)^2 (84) (06) (26)^3 (46)^2 (66) (08) (28) (48)
[633]	(20) (60) (02) (22) (42)^2 (62) (24)^2 (44)^2 (64) (06) (26) (46) (66)

[m]	$(2S, 2T)^{d([m]ST)}$
[1]	(11)
[3]	(11) (33)
[21]	(11) (31) (13)
[5]	(11) (33) (55)
[41]	(11) (31) (13) (33) (53) (35)
[32]	(11) (31) (51) (13) (33) (15)
[311]	(11) (31) (13) (33)
[61]	(11) (31) (13) (33) (53) (35) (55) (75) (57)
[52]	(11) (31) (51) (13) (33)^2 (53) (73) (15) (35) (55) (37)
[511]	(11) (31) (13) (33) (53) (35) (55)
[43]	(11) (31) (51) (71) (13) (33) (53) (15) (35) (17)
[421]	(11) (31)^2 (51) (13)^2 (33)^2 (53) (15) (35)
[63]	(11) (31) (51) (71) (13) (33)^2 (53)^2 (73) (93) (15) (35)^2 (55) (75) (17) (37) (57) (39)
[621]	(11) (31)^2 (51) (13)^2 (33)^2 (53)^2 (73) (15) (35)^2 (55)^2 (75) (37) (57)
[54]	(11) (31) (51) (71) (91) (13) (33) (53) (73) (15) (35) (55) (17) (37) (19)
[531]	(11) (31)^2 (51)^2 (71) (13)^2 (33)^3 (53)^2 (73) (15)^2(35)^2(55) (17) (37)
[522]	(11) (31) (51) (13) (33)^2 (53) (15) (35) (55)
[65]	(11) (31) (51) (71) (91) (11,1) (13) (33) (53) (73) (93) (15) (35)(55)(75) (17) (37) (57) (19) (39) (1,11)
[641]	(11) (31)^2 (51)^2 (71)^2 (91) (13)^2 (33)^3 (53)^3 (73)^2 (93) (15)^2 (35)^3 (55)^2 (75) (17)^2 (37)^2 (57) (19) (39)
[632]	(11) (31)^2 (51)^2 (71) (13)^2 (33)^3 (53)^3 (73) (15)^2 (35)^3 (55)^2 (75) (17) (37) (57)

We can get the missing labels ω and φ if we are able to diagonalize Ω and Φ in the basis states of an IR of SU(4), characterized by given values of S, T, M_S, and M_T:

$$\Omega\,|\,[m_1 m_2 m_3]\omega\varphi STM_SM_T\rangle = \omega\,|\,[m_1 m_2 m_3]\omega\varphi STM_SM_T\rangle,$$
(5.4)

$$\Phi\,|\,[m_1 m_2 m_3]\omega\varphi STM_SM_T\rangle = \varphi\,|\,[m_1 m_2 m_3]\omega\varphi STM_SM_T\rangle.$$
(5.5)

They are, of course, independent of M_S and M_T.

From the theory developed in Sec. 4, we can obtain the matrices of Ω and Φ in the Gel'fand and Tseitlin basis. By rearranging their rows and columns if necessary, these matrices can be put in a block-diagonal form, each block being characterized by given values of M_S and M_T [see Eqs. (4.4a) and (4.4b)].

We first consider the block corresponding to the highest possible value of M_S,

$$P = \tfrac{1}{2}(m_1 + m_2 - m_3),$$
(5.6)

and to the highest value of M_T compatible with this value of M_S,

$$P' = \tfrac{1}{2}(m_1 - m_2 + m_3).$$
(5.7)

It corresponds to the highest weight state of the IR $[m_1 m_2 m_3]$ of SU(4), and it is therefore one-dimensional. We thus obtain easily the eigenvalues of Ω and Φ in the state characterized by $S = P$ and $T = P'$.

We then consider the block corresponding to $M_S = P$ and $M_T = P' - 1$, and diagonalize it. The eigenvalues we get are associated with states with $S = P$ and $T = P'$ or $P' - 1$. As we already know the eigenvalue corresponding to the highest weight state, we can see at once which eigenvalues correspond to the states with $S = P$ and $T = P' - 1$.

We then consider the block corresponding to $M_S = P$ and $M_T = P' - 2$, and proceed in the same way. When we

have diagonalized all the blocks with $M_S = P$ and $M_T \geq 0$, we consider those corresponding to $M_S = P - 1$ and all nonnegative values of M_T, starting from the highest one. Thus by diagonalizing successively all the blocks with $M_S \geq 0$ and $M_T \geq 0$, we get the eigenvalues of Ω and Φ and the expansions of the corresponding eigenvectors in the Gel'fand and Tseitlin basis:

$$
|\,[m_1 m_2 m_3]\omega\varphi STM_SM_T\rangle
$$
$$
= \sideset{}{'}\sum_{m_{ij}} a([m_1 m_2 m_3]m_{ij}\omega\varphi ST)
$$
$$
\times \left|\begin{matrix} m_1 & m_2 & m_3 & 0 \\ & m_{13} & m_{23} & m_{33} \\ & m_{12} & m_{22} & \\ & m_{11} & & \end{matrix}\right\rangle.
$$
(5.8)

Here the prime on the summation symbol means that the values of m_{ij}, $1 \leq i$, $j \leq 3$, satisfy Eqs. (4.4a) and (4.4b).

The procedure described above is somewhat heavy and redundant, but it has the advantage of producing a simple method for checking the eigenvalues. Numerical calculations were performed for all the IR's of SU(4) appearing in light nuclei up to the $s-d$ shell. Tables III and IV list the eigenvalues ω and φ for those cases where the multiplicity d ($[m_1 m_2 m_3]$ ST) is greater than 1. Those corresponding to a multiplicity equal to 1 can be deduced from Tables V–X. In reading the tables one has to take into account that ω and φ are invariant under a permutation of S and T.

Let us quote some cases where the eigenvalues ω and φ are equal to zero. First of all, from the definition of Ω and Φ, it is clear that

$$\omega = 0 \quad \text{whenever} \quad S \text{ or } T = 0,$$
(5.9)

and

$$\varphi = 0 \quad \text{whenever} \quad S = T = 0.$$
(5.10)

On the other hand, the symmetry relation (4.6) shows that

TABLE III. Eigenvalues of Ω and Φ corresponding to IR's [m] of SU(4) with $\sum_i m_i$ even. For each set of labels [m], S and T, the first row contains all possible values of 16ω, and the second one the corresponding values of 64φ.

[m]	S	T	Eigenvalues			
[42]	1	1	− 5.612497	45.612497	—	—
			− 304.249939	720.249939	—	—
[321]	1	1	− 16	16	—	—
			48	48	—	—
[62]	1	1	− 12.660606	60.660606	—	—
			− 239.854533	1519.854533	—	—
	2	2	103.337900	200.662100	—	—
			1904.109594	4239.890406	—	—
[53]	2	1	− 12.660606	60.660606	—	—
			− 815.854533	943.854533	—	—
[521]	1	1	− 24	24	—	—
			208	528	—	—
	2	1	11.339394	84.660606	—	—
			824.072733	1703.927267	—	—
	2	2	72	120	—	—
			1488	1296	—	—
[431]	1	1	0	0	—	—
			− 192	448	—	—
	2	1	− 27.712813	27.712813	—	—
			− 128	− 128	—	—
[422]	1	1	− 39.191836	39.191836	—	—
			448	448	—	—
[64]	1	1	− 4.249031	60.249031	—	—
			− 310.972868	1494.972868	—	—
	3	1	− 20.166378	76.166378	—	—
			− 1556.658593	1140.658593	—	—
	2	2	− 27.569776	83.569776	—	—
			− 2115.953727	995.953727	—	—
[631]	1	1	− 8	8	—	—
			− 160	992	—	—
	2	1	− 42.973727	38.370868	100.602859	—
			92.000387	680.354199	2619.645414	—
	3	1	14.568733	105.431267	—	—
			1230.824799	2321.175201	—	—
	2	2	8	96	120	—
			864	2624	1760	—
	3	2	84.604569	163.395431	—	—
			1805.581725	1490.418275	—	—
[622]	1	1	− 46.754310	54.754310	—	—
			772.982759	1179.017241	—	—
	2	2	26.569761	173.430239	—	—
			2218.279044	2805.720956	—	—
[541]	1	1	− 16	16	—	—
			368	368	—	—
	2	1	0	0	—	—
			− 748.179617	588.179617	—	—
	3	1	− 39.191836	39.191836	—	—
			− 432	− 432	—	—
	2	2	− 48	48	—	—
			− 784	− 784	—	—
[532]	1	1	− 40	40	—	—
			608	608	—	—
	2	1	− 60.398675	0	60.398675	—
			672	1280	672	—
	2	2	− 24	24	—	—
			992	992	—	—
[651]	1	1	0	0	—	—
			64	960	—	—
	2	1	− 27.712813	27.712813	—	—
			256	256	—	—
	3	1	0	0	—	—
			− 1556.974301	660.974301	—	—
	4	1	− 50.596443	50.596443	—	—
			− 864	− 864	—	—
	2	2	0	0	—	—
			− 2112	576	—	—
	3	2	− 67.882251	67.882251	—	—
			− 1824	− 1824	—	—
[642]	2	0	0	0	—	—
			71.527669	2104.472331	—	—

1459   J. Math. Phys., Vol. 17, No. 8, August 1976

C. Quesne   1459

TABLE III. (Continued).

	1	1	− 53.065997	0	53.065997	—
			1088	− 320	1088	—
	2	1	− 53.065997	0	53.065997	—
			640	2048	640	—
	3	1	− 81.584312	0	81.584312	—
			928	1760	928	—
	2	2	− 99.919968	.0	0	99.919968
			448	1216	2112	448
	3	2	− 39.191836	39.191836	—	—
			1120	1120	—	—
[633]	2	1	− 87.635609	87.635609	—	—
			1792	1792	—	—
	2	2	− 67.882251	67.882251	—	—
			2112	2112	—	—

TABLE IV. Eigenvalues of Ω and Φ corresponding to IR's [m] of SU(4) with $\sum_i m_i$ odd. For each set of labels [m], S and T, the first row contains all possible values of 16ω, and the second one the corresponding values of 64φ.

[m]	S	T		Eigenvalues	
[52]	$\frac{3}{2}$	$\frac{3}{2}$	36.128808	105.871192	—
			434.833786	1969.166214	—
[421]	$\frac{3}{2}$	$\frac{1}{2}$	− 19.083189	29.083189	—
			145.168108	626.831892	—
	$\frac{3}{2}$	$\frac{3}{2}$	15	55	—
			498	418	—
[63]	$\frac{3}{2}$	$\frac{3}{2}$	− 22.105555	120.105555	—
			− 934.744439	2762.744439	—
	$\frac{5}{2}$	$\frac{3}{2}$	45	133	—
			270	2558	—
[621]	$\frac{3}{2}$	$\frac{1}{2}$	− 24.048349	38.048349	—
			431.323109	1300.676891	—
	$\frac{3}{2}$	$\frac{3}{2}$	21	77	—
			1362	1666	—
	$\frac{5}{2}$	$\frac{3}{2}$	65.297114	172.702886	—
			2162.159591	3665.840409	—
	$\frac{5}{2}$	$\frac{5}{2}$	161	217	—
			3282	2946	—
[531]	$\frac{3}{2}$	$\frac{1}{2}$	− 5.493902	35.493902	—
			− 148.817046	1080.817046	—
	$\frac{5}{2}$	$\frac{1}{2}$	− 27.249031	37.249031	—
			183.509690	828.490310	—
	$\frac{3}{2}$	$\frac{3}{2}$	− 39.660606	33.660606	45
			− 145.890900	1173.890900	418
	$\frac{5}{2}$	$\frac{3}{2}$	13.473891	80.526109	—
			501.052218	366.947782	—
[522]	$\frac{3}{2}$	$\frac{3}{2}$	− 17.258640	91.258640	—
			1045.482720	1262.517280	—
[641]	$\frac{3}{2}$	$\frac{1}{2}$	− 22.559468	40.559468	—
			377.929582	1514.070418	—
	$\frac{5}{2}$	$\frac{1}{2}$	− 6.612497	44.612497	—
			− 467.274884	1479.274884	—
	$\frac{7}{2}$	$\frac{1}{2}$	− 35.447497	45.447497	—
			213.525032	1022.474968	—
	$\frac{3}{2}$	$\frac{3}{2}$	− 9.660606	27	63.660606
			− 969.818167	1474	1229.818167
	$\frac{5}{2}$	$\frac{3}{2}$	− 60.588290	42.921131	58.667160
			− 678.294081	1564.461078	27.833003
	$\frac{7}{2}$	$\frac{3}{2}$	12.352385	105.647615	—
			399.295230	212.704770	—
	$\frac{5}{2}$	$\frac{5}{2}$	7	119	—
			− 46	− 270	—
[632]	$\frac{3}{2}$	$\frac{1}{2}$	− 34.559468	28.559468	—
			1039.356806	660.643194	—
	$\frac{5}{2}$	$\frac{1}{2}$	− 43.754310	57.754310	—
			1379.439658	2800.560342	—
	$\frac{3}{2}$	$\frac{3}{2}$	− 72.635609	− 9	102.635609
			804.186345	1378	1855.813655
	$\frac{5}{2}$	$\frac{3}{2}$	− 25.808121	56.686872	130.121249
			1616.918762	2572.932601	1788.148637
	$\frac{5}{2}$	$\frac{5}{2}$	35	91	—
			2130	2018	—

$$\langle\!\langle C^{(111)}\rangle\!\rangle^{[m_1\, m_1-m_3\, m_1-m_2]-M_S-M}T = -\langle\!\langle C^{(111)}\rangle\!\rangle^{[m_1 m_2 m_3]M_S M}T,$$

(5.11)

i. e., the trace of $C^{(111)}$ in the states of the IR $[m_1\ m_1 - m_3\ m_1 - m_2]$, characterized by $- M_S$ and $- M_T$, is equal and opposite to that of $C^{(111)}$ in the states of the IR $[m_1 m_2 m_3]$, specified by M_S and M_T. Therefore, for any self-contragredient IR, i. e., such that $m_1 = m_2 + m_3$, the trace of $C^{(111)}$ in the states corresponding to fixed values of M_S and M_T is zero. In other words, for a self-contragredient representation, the sum of the eigenvalues ω corresponding to fixed values of S and T is zero. In particular,

$$\omega = 0 \quad \text{whenever} \quad m_1 = m_2 + m_3 \quad \text{and} \quad d([m_1\, m_2\, m_3]ST) = 1.$$

(5.12)

We also have explicit formulas for the eigenvalues ω and φ in three special cases. For the highest weight state of any IR $[m_1 m_2 m_3]$, characterized by values of S and T given by Eqs. (5.6) and (5.7) respectively, we get

$$\omega = \tfrac{1}{2}(P + 1)(P' + 1)P''$$

(5.13)

and

$$\varphi = \tfrac{1}{4}(P + 1)[(P + 1)(P''^2 + P') - 3P]$$
$$+ \tfrac{1}{4}(P' + 1)[(P' + 1)P''^2 + 3PP'],$$

(5.14)

where

$$P'' = \tfrac{1}{2}(m_1 - m_2 - m_3).$$

(5.15)

For the state of the IR $[m\ 0\ 0]$ with $S = T = m/2 - a$, where a is any integer such that $0 \leq a \leq [m/2]$, ω and φ are given by

$$\omega = \tfrac{1}{16}(m + 2)(m - 2a)(m - 2a + 2),$$

(5.16)

and

$$\varphi = \tfrac{1}{32}(m - 2a)(m - 2a + 2)[m^2 + (4a + 6)m - 4a^2 + 4a - 4].$$

(5.17)

Finally, for the state of the IR $[m\ m\ 0]$ with $S = m - a$ and $T = a$, where a is any integer such that $0 \leq a \leq m$, we have

$$\varphi = - \tfrac{1}{4}[(2a^2 + 2a + 3)m^2 + (- 4a^3 - 2a^2 - 4a + 3)m + 2a^2(a^2 + 2)],$$

(5.18)

and $\omega = 0$ owing to Eq. (5.12).

TABLE V. Averages $\langle 16\ C^{(111)}\rangle^{[m]S}_T$ corresponding to IR's [m] of U(4) with $\sum_i m_i$ even.

[m]	S T 11	21	31	41	51	22	32	42	33
[2]	32	—	—	—	—	—	—	—	—
[4]	48	—	—	—	—	144	—	—	—
[31]	32	48	—	—	—	—	—	—	—
[6]	64	—	—	—	—	192	—	—	384
[51]	40	72	—	—	—	168	192	—	—
[42]	20	40	64	—	—	72	—	—	—
[411]	8	72	—	—	—	72	—	—	—
[62]	24	48	96	—	—	152	208	240	256
[611]	16	96	—	—	—	112	256	—	256
[53]	48	24	48	80	—	48	96	—	—
[521]	0	48	96	—	—	96	96	—	—
[64]	28	56	28	56	96	28	56	120	128
[631]	0	32	60	120	—	224/3	124	120	128
[622]	4	8	128	—	—	100	128	—	128

TABLE VI. Averages $\langle 16\ C^{(111)}\rangle^{[m]S}_T$ corresponding to IR's [m] of U(4) with $\sum_i m_i$ odd.

[m]	S T $\frac{1}{2}\frac{1}{2}$	$\frac{3}{2}\frac{1}{2}$	$\frac{5}{2}\frac{1}{2}$	$\frac{7}{2}\frac{1}{2}$	$\frac{9}{2}\frac{1}{2}$	$\frac{11}{2}\frac{1}{2}$	$\frac{3}{2}\frac{3}{2}$	$\frac{5}{2}\frac{3}{2}$	$\frac{7}{2}\frac{3}{2}$	$\frac{9}{2}\frac{3}{2}$	$\frac{5}{2}\frac{5}{2}$	$\frac{7}{2}\frac{5}{2}$
[1]	9	—	—	—	—	—	—	—	—	—	—	—
[3]	15	—	—	—	—	—	75	—	—	—	—	—
[21]	3	15	—	—	—	—	—	—	—	—	—	—
[5]	21	—	—	—	—	—	105	—	—	—	245	—
[41]	1	25	—	—	—	—	85	105	—	—	—	—
[32]	13	1	21	—	—	—	25	—	—	—	—	—
[311]	−11	25	—	—	—	—	25	—	—	—	—	—
[61]	−1	35	—	—	—	—	107	147	—	—	287	315
[52]	19	−5	35	—	—	—	71	107	135	—	147	—
[511]	−13	35	—	—	—	—	47	147	—	—	147	—
[43]	7	19	−1	27	—	—	−5	35	—	—	—	—
[421]	−5	5	35	—	—	—	35	35	—	—	—	—
[63]	5	29	−11	45	—	—	49	89	129	165	133	189
[621]	−7	7	49	—	—	—	49	119	189	—	189	189
[54]	17	5	25	−3	33	—	29	−11	45	—	49	—
[531]	−15	15	5	45	—	—	13	47	45	—	49	—
[522]	17	−31	49	—	—	—	37	49	—	—	49	—
[65]	11	23	3	31	−5	39	−1	39	−17	55	−21	63
[641]	−9	9	19	5	55	—	27	41/3	59	55	63	63
[632]	3	−3	7	63	—	—	7	161/3	63	—	63	63

TABLE VII. Averages $\langle 64\ C^{(202)}\rangle^{[m]S}_T$ corresponding to IR's [m] of U(4) with $\sum_i m_i$ even. (a) $T = 0, 1$. (b) $T = 2, 3, 4, 5$.

(a) [m]	S T 10	20	30	40	50	60	11	21	31	41	51
[2]	—	—	—	—	—	—	32	—	—	—	—
[11]	−96	—	—	—	—	—	—	—	—	—	—
[4]	—	—	—	—	—	—	192	—	—	—	—
[31]	−32	—	—	—	—	—	160	0	—	—	—
[22]	—	−288	—	—	—	—	−96	—	—	—	—
[211]	96	—	—	—	—	—	−32	—	—	—	—
[6]	—	—	—	—	—	—	416	—	—	—	—
[51]	96	—	—	—	—	—	368	432	—	—	—
[42]	—	−96	—	—	—	—	136	176	−64	—	—
[411]	288	—	—	—	—	—	80	720	—	—	—
[33]	−96	—	−576	—	—	—	—	−288	—	—	—
[321]	144	144	—	—	—	—	56	−144	—	—	—
[62]	—	288	—	—	—	—	320	672	768	—	—
[611]	544	—	—	—	—	—	256	1440	—	—	—
[53]	−32	—	−192	—	—	—	480	128	160	−160	—
[521]	304	624	—	—	—	—	216	728	1216	—	—
[44]	—	−288	—	−960	—	—	−96	—	−576	—	—
[431]	352	96	192	—	—	—	96	32	−320	—	—
[422]	—	864	—	—	—	—	176	576	—	—	—
[64]	—	−96	—	−320	—	—	328	560	88	112	−288
[631]	608	416	1088	—	—	—	240	960	1080	1840	—
[622]	—	1632	—	—	—	—	376	1232	3008	—	—
[55]	−96	—	−576	—	−1440	—	—	−288	—	−960	—

TABLE VII. (*Continued*)

[m]											
[541]	464	464	-16	240	—	—	216	24	-24	-560	—
[532]	32	1376	1472	—	—	—	336	640	960	—	—
[66]	—	-288	—	-960	—	-2016	-96	—	-576	—	-1440
[651]	736	480	576	-192	288	—	288	224	-112	-112	-864
[642]	—	1088	1984	2240	—	—	352	992	992	1440	—
[633]	544	—	3264	—	—	—	192	1424	2752	—	—

(b) [m] \ S T	12	22	32	42	13	23	33	14	24	15
[2]	—	—	—	—	—	—	—	—	—	—
[11]	—	—	—	—	—	—	—	—	—	—
[4]	—	576	—	—	—	—	—	—	—	—
[31]	64	—	—	—	—	—	—	—	—	—
[22]	—	—	—	—	—	—	—	—	—	—
[211]	—	—	—	—	—	—	—	—	—	—
[6]	—	1248	—	—	—	—	2496	—	—	—
[51]	240	1200	960	—	—	672	—	—	—	—
[42]	240	48	—	—	96	—	—	—	—	—
[411]	144	144	—	—	—	—	—	—	—	—
[33]	-96	—	—	—	—	—	—	—	—	—
[321]	0	—	—	—	—	—	—	—	—	—
[62]	480	1216	1760	1440	288	1472	1088	—	768	—
[611]	352	640	2752	—	—	1120	1216	—	—	—
[53]	192	288	0	—	320	96	—	128	—	—
[521]	216	600	192	—	176	240	—	—	—	—
[44]	—	-288	—	—	-96	—	—	—	—	—
[431]	96	-96	—	—	32	—	—	—	—	—
[422]	0	0	—	—	—	—	—	—	—	—
[64]	624	200	304	-80	248	400	64	400	144	160
[631]	1024/3	2368/3	920	240	280	760	320	208	336	—
[622]	144	824	1984	—	288	352	448	—	—	—
[55]	-96	—	-576	—	—	-288	—	-96	—	—
[541]	152	88	-256	—	136	-48	—	64	—	—
[532]	448/3	400	-64	—	32	96	—	—	—	—
[66]	—	-288	—	-960	-96	—	-576	—	-288	-96
[651]	288	96	48	-480	208	144	-192	176	0	96
[642]	288	608	592	-160	192	560	64	64	192	—
[633]	336	624	1728	—	32	96	192	—	—	—

TABLE VIII. Averages $\langle 64\, C^{(202)}\rangle^{[m]ST}$ corresponding to IR's [m] of U(4) with $\sum_i m_i$ odd. (a) $T=\frac{1}{2}, \frac{3}{2}$. (b) $T=\frac{5}{2}, \frac{7}{2}, \frac{9}{2}, \frac{11}{2}$.

(a) [m] \ S T	$\frac{1}{2}\frac{1}{2}$	$\frac{3}{2}\frac{1}{2}$	$\frac{5}{2}\frac{1}{2}$	$\frac{7}{2}\frac{1}{2}$	$\frac{9}{2}\frac{1}{2}$	$\frac{11}{2}\frac{1}{2}$	$\frac{1}{2}\frac{3}{2}$	$\frac{3}{2}\frac{3}{2}$	$\frac{5}{2}\frac{3}{2}$	$\frac{7}{2}\frac{3}{2}$	$\frac{9}{2}\frac{3}{2}$
[1]	-9	—	—	—	—	—	—	—	—	—	—
[3]	39	—	—	—	—	—	—	195	—	—	—
[21]	15	-105	—	—	—	—	3	—	—	—	—
[5]	111	—	—	—	—	—	—	555	—	—	—
[41]	71	95	—	—	—	—	59	515	315	—	—
[32]	47	-73	-273	—	—	—	35	-85	—	—	—
[311]	31	295	—	—	—	—	19	-5	—	—	—
[61]	151	415	—	—	—	—	139	979	1099	—	—
[52]	111	135	175	—	—	—	99	507	739	459	—
[511]	95	695	—	—	—	—	83	251	1491	—	—
[43]	87	-33	-233	-513	—	—	75	-45	-245	—	—
[421]	63	303	511	—	—	—	51	219	-77	—	—
[63]	159	183	223	279	—	—	147	747	707	987	627
[621]	135	663	1351	—	—	—	123	579	1499	2403	—
[54]	135	15	-185	-465	-825	—	123	3	-197	-477	—
[531]	103	503	423	783	—	—	91	905/3	315	-189	—
[522]	87	687	1687	—	—	—	75	387	1099	—	—
[65]	191	71	-129	-409	-769	-1209	179	59	-141	-421	-781
[641]	151	743	663	551	1111	—	139	1433/3	1153/3	419	-341
[632]	127	623	1983	2727	—	—	115	2033/3	3673/3	1755	—

(b) [m] \ S T	$\frac{1}{2}\frac{5}{2}$	$\frac{3}{2}\frac{5}{2}$	$\frac{5}{2}\frac{5}{2}$	$\frac{7}{2}\frac{5}{2}$	$\frac{1}{2}\frac{7}{2}$	$\frac{3}{2}\frac{7}{2}$	$\frac{5}{2}\frac{7}{2}$	$\frac{1}{2}\frac{9}{2}$	$\frac{3}{2}\frac{9}{2}$	$\frac{1}{2}\frac{11}{2}$
[1]	—	—	—	—	—	—	—	—	—	—

[m]										
[3]	—	—	—	—	—	—	—	—	—	—
[21]	—	—	—	—	—	—	—	—	—	—
[5]	—	—	1295	—	—	—	—	—	—	—
[41]	—	255	—	—	—	—	—	—	—	—
[32]	15	—	—	—	—	—	—	—	—	—
[311]	—	—	—	—	—	—	—	—	—	—
[61]	—	639	2359	2079	—	—	1435	—	—	—
[52]	79	679	399	—	—	315	—	—	—	—
[511]	—	471	511	—	—	—	—	—	—	—
[43]	55	−65	—	—	27	—	—	—	—	—
[421]	31	55	—	—	—	—	—	—	—	—
[63]	127	647	959	567	99	843	483	—	375	—
[621]	103	599	1319	783	—	531	651	—	—	—
[54]	103	−17	−217	—	75	−45	—	39	—	—
[531]	71	311	7	—	43	115	—	—	—	—
[522]	55	79	119	—	—	—	—	—	—	—
[65]	159	39	−161	−441	131	11	−189	95	−25	51
[641]	119	1213/3	439	−81	91	403	91	55	175	—
[632]	95	1093/3	799	135	67	139	259	—	—	—

TABLE IX. Averages $\langle 64\, C^{(112)}\rangle^{[\bar{m}]ST}$ corresponding to IR's [m] of U(4) with $\sum_i m_i$ even.

[m] $\diagdown$ ST	11	21	31	41	51	22	32	42	33
[2]	−128	—	—	—	—	—	—	—	—
[4]	−448	—	—	—	—	−576	—	—	—
[31]	0	−192	—	—	—	—	—	—	—
[22]	128	—	—	—	—	—	—	—	—
[211]	−128	—	—	—	—	—	—	—	—
[6]	−896	—	—	—	—	−1920	—	—	−1536
[51]	−128	−768	—	—	—	−384	−768	—	—
[42]	64	128	−256	—	—	0	—	—	—
[411]	−320	−192	—	—	—	−576	—	—	—
[33]	—	384	—	—	—	—	—	—	—
[321]	64	−192	—	—	—	—	—	—	—
[62]	0	0	−1152	—	—	−640	−320	−960	−512
[611]	−576	−768	—	—	—	−1728	−768	—	−1536
[53]	0	256	320	−320	—	768	384	—	—
[521]	64	−320	−256	—	—	−192	−768	—	—
[44]	128	—	768	—	—	1152	—	—	—
[431]	64	256	−256	—	—	0	—	—	—
[422]	−96	−192	—	—	—	−576	—	—	—
[64]	64	128	544	576	−384	960	1728	960	1536
[631]	64	512/3	−416	−320	—	−512/3	32	−960	−512
[622]	−224	−448	−256	—	—	−864	−768	—	−1536
[55]	—	384	—	1280	—	—	2304	—	—
[541]	64	256	544	−320	—	960	384	—	—
[532]	64	−256/3	−256	—	—	−192	−768	—	—
[66]	128	—	768	—	1920	1152	—	3840	4608
[651]	64	256	544	928	−384	960	2016	960	1536
[642]	256/3	512/3	−64/3	−320	—	160	32	−960	−512
[633]	−576	−32	−256	—	—	−864	−768	—	−1536

TABLE X. Averages $\langle 64\, C^{(112)}\rangle^{[\bar{m}]ST}$ corresponding to IR's [m] of U(4) with $\sum_i m_i$ odd.

[m] $\diagdown$ ST	$\frac{1}{2}\frac{1}{2}$	$\frac{3}{2}\frac{1}{2}$	$\frac{5}{2}\frac{1}{2}$	$\frac{7}{2}\frac{1}{2}$	$\frac{9}{2}\frac{1}{2}$	$\frac{11}{2}\frac{1}{2}$	$\frac{3}{2}\frac{3}{2}$	$\frac{5}{2}\frac{3}{2}$	$\frac{7}{2}\frac{3}{2}$	$\frac{9}{2}\frac{3}{2}$	$\frac{5}{2}\frac{5}{2}$	$\frac{7}{2}\frac{5}{2}$
[1]	−36	—	—	—	—	—	—	—	—	—	—	—
[3]	−132	—	—	—	—	—	−300	—	—	—	—	—
[21]	60	−60	—	—	—	—	—	—	—	—	—	—
[5]	−276	—	—	—	—	—	−1020	—	—	—	−980	—
[41]	124	−260	—	—	—	—	−140	−420	—	—	—	—
[32]	−68	196	−84	—	—	—	100	—	—	—	—	—
[311]	−36	−60	—	—	—	—	−300	—	—	—	—	—
[61]	204	−540	—	—	—	—	−636	−1596	—	—	−756	−1260
[52]	−196	380	−420	—	—	—	−188	−36	−540	—	−196	—
[511]	−20	−260	—	—	—	—	−860	−420	—	—	−980	—
[43]	92	−124	396	−108	—	—	740	420	—	—	—	—
[421]	92	−32	−84	—	—	—	−20	−420	—	—	—	—
[63]	188	−388	732	−612	—	—	580	−60	132	−660	700	252
[621]	140	−80	−420	—	—	—	−356	−816	−540	—	−476	−1260
[54]	−100	260	−180	660	−132	—	−28	1572	900	—	1372	—

TABLE X. (Continued) ·

[531]	-68	128	-12	-108	—	—	364/3	192	-540	—	-196	—
[522]	-148	-4	-84	—	—	—	-380	-420	—	—	-980	—
[65]	124	-188	492	-236	988	-156	868	228	2692	1540	3388	2772
[641]	124	-64	276	24	-132	—	1132/3	484	516	-660	1036	252
[632]	172	-112	-12	-108	—	—	28/3	-404	-540	—	-476	-1260

TABLE XI. Coefficients $a_{\rho'r}$ in fractional form $N_{\rho'r}/D$. The numerators $N_{\rho'r}$ are given in the table and the common denominator is $D = 1152$.

Each ρ' row has two sub-rows: the upper corresponds to $r = 1$–14, the lower to $r = 15$–28.

ρ' \ r	1 / 15	2 / 16	3 / 17	4 / 18	5 / 19	6 / 20	7 / 21	8 / 22	9 / 23	10 / 24	11 / 25	12 / 26	13 / 27	14 / 28
1	1152	-2400	1680	-480	48	—	—	—	—	—	—	—	—	—
	—	—	—	—	—	—	—	—	—	—	—	—	—	—
2	—	4608	-4992	1728	-192	—	—	—	—	—	—	—	—	—
	—	—	—	—	—	—	—	—	—	—	—	—	—	—
3	—	-7776	6264	-1656	144	864	-504	72	—	—	—	—	1728	-1008
	144	1728	-1008	144	—	—	—	—	—	—	—	—	—	—
4	—	-864	504	-72	—	864	-504	72	—	—	—	—	-1728	1008
	-144	-1728	1008	-144	—	—	—	—	—	—	—	—	—	—
5	—	2592	-648	-288	72	-864	504	-72	—	—	—	—	1728	-1008
	144	-1728	1008	-144	—	—	—	—	—	—	—	—	—	—
6	—	2592	-648	-288	72	-864	504	-72	—	—	—	—	-1728	1008
	-144	1728	-1008	144	—	—	—	—	—	—	—	—	—	—
7	—	6144	-3584	640	-32	-1024	512	-64	—	—	—	—	-2048	1024
	-128	-2048	1024	-128	—	—	—	—	—	2048	-512	—	—	—
8	—	1536	-512	32	—	-1792	576	-32	256	-64	—	—	2048	-1024
	128	2048	-1024	128	—	—	—	—	—	-2048	512	—	—	—
9	—	-3072	512	320	-64	1024	-512	64	—	—	—	—	-1024	1280
	-256	2048	-1024	128	—	—	—	—	—	-2048	512	—	—	—
10	—	-3072	512	320	-64	1024	-512	64	—	—	—	—	2048	-1024
	128	-1024	1280	-256	—	—	—	—	—	-2048	512	—	—	—
11	—	-1536	-128	128	—	2048	—	-128	-512	128	—	—	-1024	-256
	128	-1024	-256	128	—	—	—	—	—	4096	-1024	—	—	—
12	—	1536	512	-96	-32	-1280	-64	96	256	-64	—	—	—	—
	—	—	—	—	—	—	—	—	—	—	—	—	—	—
13	—	-1800	738	-73	2	348	-162	9	16	4	-4	4	776	-300
	24	776	-300	24	8	40	8	4	4	-1200	200	96	96	-16
14	—	-648	90	-3	—	828	-102	3	-192	12	12	—	-744	252
	-24	-744	252	-24	-24	-72	-24	12	12	1008	-168	-96	-96	-48
15	—	-72	6	—	—	108	-8	—	-40	—	4	2	-32	48
	—	-32	48	—	16	32	16	-16	-16	192	-32	—	—	64
16	—	1080	-54	-111	9	-324	186	3	-48	-12	12	-12	120	-420
	72	-888	300	-12	24	-24	-24	12	-12	1200	-120	-96	-288	48
17	—	1080	-54	-111	9	-324	186	3	-48	-12	12	-12	-888	300
	-12	120	-420	72	-24	-24	24	-12	12	1200	-120	-288	-96	48
18	—	648	162	-18	—	-972	-180	18	360	—	-36	18	504	144
	-18	504	144	-18	—	—	—	-36	-36	-1440	288	288	288	—
19	—	216	54	-6	—	-324	-60	6	120	12	-12	-6	96	-48
	-24	168	24	-18	-48	24	48	48	-12	-480	-24	—	96	-48
20	—	216	54	-6	—	-324	-60	6	120	12	-12	-6	168	24
	-18	96	-48	-24	48	24	-48	-12	48	-480	-24	96	—	-48
21	—	—	-24	4	4	-24	-24	-12	32	8	-8	8	-32	—
	24	112	—	-12	16	-16	16	-16	8	—	-80	—	192	-32
22	—	—	-72	-24	—	72	120	24	-96	-48	24	—	-48	—
	-12	-48	—	-12	-48	—	-48	24	24	—	96	-192	-192	192
23	—	—	-24	4	4	-24	-24	-12	32	8	-8	8	112	—
	-12	-32	—	24	16	-16	16	8	-16	—	-80	192	—	-32
24	—	—	—	—	—	-24	-8	—	32	—	-8	8	-32	—
	—	-32	—	—	16	32	16	-16	-16	—	64	—	—	-128
25	—	-648	-162	72	18	324	-144	-72	72	72	-36	18	216	72
	18	216	72	18	—	144	—	-36	-36	-576	-144	288	288	-288
26	—	-216	-126	-18	—	396	180	18	-216	-36	36	-18	—	—
	—	-216	-72	-18	—	-72	—	—	36	288	72	—	-288	144
27	—	-216	-126	-18	—	396	180	18	-216	-36	36	-18	-216	-72
	-18	—	—	—	—	-72	—	36	—	288	72	-288	—	144
28	—	72	66	20	2	-132	-80	-12	—	72	16	-12	6	—
	—	—	—	—	—	—	—	—	—	—	—	—	—	—

6. APPLICATION TO THE PROPAGATION OF OPERATOR AVERAGES

In Ref. 10, it is shown that the construction of an integrity basis for the invariants with respect to a subgroup that are contained in the enveloping algebra of a group is quite relevant for the propagation of operator averages,

which is an essential concept of the so-called French's spectral distribution method in nuclear spectroscopy.[9] We now proceed to illustrate this point for the case $SU(4) \subset SU(2) \times SU(2)$.

Let us assume that n nucleons are distributed over some finite set of $\mathcal{N}$ single-particle states, and let us separate the spatial coordinates of these nucleons from their spin and isospin coordinates. The n-particle states can be classified according to the IR's of the chain of groups

$$\underset{[1^n]}{U(\mathcal{N})} \supset \underset{[\tilde{m}]}{U(\mathcal{N}/4)} \times \underset{[\tilde{m}]}{U(4)}, \qquad (6.1)$$

where $U(\mathcal{N}/4)$ is the group of unitary transformations acting on the spatial part of the single-particle states, and $U(4)$ is Wigner's supermultiplet group, containing the spin and isospin $SU(2)$ subgroups,

$$\underset{[m]}{U(4)} \supset \underset{S}{SU(2)} \times \underset{T}{SU(2)} \supset \underset{M_S}{U(1)} \times \underset{M_T}{U(1)}. \qquad (6.2)$$

The n-particle states are denoted by $|[\tilde{m}]\mu;$ $[m]\omega\varphi STM_SM_T\rangle$, where $[\tilde{m}]$ is obtained from $[m]$ by changing rows into columns and μ specifies the row of the IR $[\tilde{m}]$ of $U(\mathcal{N}/4)$.

When considering the moments of the fixed supermultiplet (or equivalently fixed spatial symmetry), spin and isospin spectral distributions,[16] one has to evaluate operator averages such as

$$\langle O \rangle^{[\tilde{m}]ST}$$

$$= \{\dim([\tilde{m}]) \times (2S+1) \times (2T+1) \times d([m]ST)\}^{-1}$$

$$\times \sum_{\mu\omega\varphi M_S M_T} \langle [\tilde{m}]\mu; [m]\omega\varphi STM_SM_T | O | [\tilde{m}]\mu;$$

$$[m]\omega\varphi STM_SM_T\rangle, \qquad (6.3)$$

where $\dim([\tilde{m}])$ is the dimension of the IR $[\tilde{m}]$ of $U(\mathcal{N}/4)$. Average (6.3) is taken in the representation space of the IR $[\tilde{m}] \times S \times T \times [1]$ of the product group $K = U(\mathcal{N}/4) \times SU(2) \times SU(2) \times U[d([m]ST)]$, where $U[d([m]ST)]$ is a term of the direct sum of unitary groups $\sum_{[m]ST} \oplus U[d([m]ST)]$, which is the trivial complementary group of $U(\mathcal{N}/4) \times SU(2) \times SU(2)$.[10] In Eq. (6.3) and in the following ones, we drop the IR $[1]$ of $U[d([m]ST)]$ because it is always the same.

Here O is some power H^p of the Hamiltonian operator, and is therefore an operator of maximum particle rank equal to $u = 2p$. The representation spaces of the IR's $[\tilde{m}'] \times S' \times T' \times [1]$ of K, which are subspaces of the $0, 1, \ldots, u$-particle spaces are called the defining subspaces for the average (6.3). They form a set, that is often called the elementary net,[9] and that we shall denote by $\mathcal{S}$.

When $d([m']S'T') = 1$ for all the defining subspaces, the average can be propagated from them, i.e., it can be expressed for any $[\tilde{m}]ST$ as a linear combination of its values in the defining subspaces[10]:

$$\langle O \rangle^{[\tilde{m}]ST} = \sum_{\{[\tilde{m}']S'T'\} \in \mathcal{S}} Q^u([\tilde{m}']S'T'; [\tilde{m}]ST)\langle O \rangle^{[\tilde{m}']S'T'}. \qquad (6.4)$$

From Tables I and II, we see that the condition of propagation is satisfied for the centroid energies $(p=1)$ and the widths $(p=2)$. We shall restrict ourselves to these two cases in the following.

The coefficients $Q^u([\tilde{m}']S'T'; [\tilde{m}]ST)$ of the linear combinations (6.4) are the averages, in the representation space of the IR $[\tilde{m}] \times S \times T \times [1]$ of K, of the so-called propagation operators $Q^u([\tilde{m}']S'T')$. These operators are scalars with respect to K, which satisfy the system of equations

$$Q^u([\tilde{m}']S'T'; [\tilde{m}'']S''T'')$$

$$= \delta_{[\tilde{m}'], [\tilde{m}'']}\delta_{S', S''}\delta_{T', T''}, \quad \{[\tilde{m}'']S''T''\} \in \mathcal{S}. \qquad (6.5)$$

In Ref. 10, it is shown that the propagation operators can be written as polynomials of degree u in the members of an integrity basis for the $SU(2) \times SU(2)$ scalars belonging to the enveloping algebra of $U(4)$, or, in other words, as polynomials of degree u in the number operator N and the members of an integrity basis for the $SU(2) \times SU(2)$ scalars in the enveloping algebra of $SU(4)$. Their averages can thus be written in the form

$$Q^u([\tilde{m}']S'T'; [\tilde{m}]ST)$$

$$= [d([m]ST)]^{-1} \sum_{\omega\varphi} \langle [m]\omega\varphi ST | Q^u([\tilde{m}']S'T') | [m]\omega\varphi ST\rangle. \qquad (6.6)$$

For $u = 2$, we see from Sec. 3 that the propagation operators are polynomials in N, S^2, T^2, and G_2, and thus can be easily constructed.[16] For $u = 4$, they are polynomials in N, S^2, T^2, G_2, G_3, G_4, $C^{(111)}$, $C^{(202)}$, $C^{(022)}$, and $C^{(112)}$, and we can thus understand why it was impossible to build the propagation operators for the widths only in terms of Casimir operators.[16] Now the theory developed in the preceding sections enables us to construct them explicitly.

Let us denote the IR's $[\tilde{m}]ST$ by a single index ρ, defined in such a way that for the defining subspaces we have the following correspondence:

$$\rho = 1, 2, \ldots, 28: \quad [0]00, \ [1]\tfrac{1}{2}\tfrac{1}{2}, \ [11]11, \ [11]00, \ [2]10,$$

$$[2]01, \ [111]\tfrac{3}{2}\tfrac{3}{2}, \ [111]\tfrac{1}{2}\tfrac{1}{2}, \ [21]\tfrac{3}{2}\tfrac{1}{2}, \ [21]\tfrac{1}{2}\tfrac{3}{2},$$

$$[21]\tfrac{1}{2}\tfrac{1}{2}, \ [3]\tfrac{1}{2}\tfrac{1}{2}, \ [1111]22, \ [1111]11, \ [1111]00,$$

$$[211]21, \ [211]12, \ [211]11, \ [211]10, \ [211]01,$$

$$[22]20, \ [22]11, \ [22]02, \ [22]00, \ [31]11, \ [31]10,$$

$$[31]01, \ [4]00. \qquad (6.7)$$

The 28 propagation operators are then written as linear combinations of 28 $SU(2) \times SU(2)$ scalars $S^{(r)}$, $r = 1, \ldots, 28$,

$$Q^4(\rho') = \sum_{r=1}^{28} a_{\rho'r}S^{(r)}, \quad \rho' \in \mathcal{S}, \qquad (6.8)$$

where

$$S^{(r)} \quad r = 1, 2, \ldots, 28 \rightarrow \mathbb{1}, N, N^2, N^3, N^4, G_2, NG_2, N^2G_2, G_3,$$

$$NG_3, G_4, G_2^2, S^2, NS^2, N^2S^2, T^2, NT^2, N^2T^2, S^4, S^2T^2, T^4,$$

$$S^2G_2, T^2G_2, C^{(111)}, NC^{(111)}, C^{(202)}, C^{(022)}, C^{(112)}. \qquad (6.9)$$

Putting Eqs. (6.8) into Eqs. (6.5), we get a system of 28 linear equations for the coefficients $a_{\rho'r}$ corresponding to any given value of $\rho' \in \mathcal{S}$:

$$\sum_{r=1}^{28} a_{\rho'r}\langle S^{(r)}\rangle^{\rho''} = \delta_{\rho'\rho''}, \quad \rho'' \in \mathcal{S}. \qquad (6.10)$$

In order to be able to write explicitly Eqs. (6.10) and

afterwards to use Eqs. (6.8), it remains to calculate the averages $\langle S^{(r)} \rangle^\rho$ for any ρ likely to appear in applications. For that purpose it is sufficient to determine the averages of the basic scalars $\mathbb{1}$, N, G_2, G_3, G_4, $\mathbf{S}^2$, $\mathbf{T}^2$, $C^{(111)}$, $C^{(202)}$, $C^{(022)}$, and $C^{(112)}$. For the first seven ones, we get straightforwardly

$$\langle \mathbb{1} \rangle^\rho = 1, \quad \langle N \rangle^\rho = \sum_{i=1}^{4} m_i,$$

$$\langle G_2 \rangle^\rho = \sum_{i=1}^{4} m_i^2 + 3m_1 + m_2 - m_3 - 3m_4,$$

$$\langle G_3 \rangle^\rho = \sum_{i=1}^{4} m_i^3 + 6m_1^2 + 3m_2^2 - 3m_4^2$$
$$- \sum_{i<j}^{4} m_i m_j + 9m_1 - m_2 - 5m_3 - 3m_4,$$

$$\langle G_4 \rangle^\rho = \sum_{i=1}^{4} m_i^4 + 9m_1^3 + 5m_2^3 + m_3^3 - 3m_4^3 \qquad (6.11)$$
$$- \sum_{i<j}^{4} (m_i^2 m_j + m_i m_j^2) + 27m_1^2 + 5m_2^2 - 5m_3^2$$
$$- 3m_4^2 - 8m_1 m_2 - 6m_1 m_3 - 4m_1 m_4 - 4m_2 m_3$$
$$- 2m_2 m_4 + 27m_1 - 11m_2 - 13m_3 - 3m_4,$$

$$\langle \mathbf{S}^2 \rangle^\rho = S(S+1), \quad \langle \mathbf{T}^2 \rangle^\rho = T(T+1).$$

For the other ones, we use the theory developed in Sec. 4 to construct their traces in the Gel'fand and Tseitlin basis states with fixed values of M_S and M_T,

$$\langle\!\langle C^{(A\,B\,C)} \rangle\!\rangle^{[\tilde{m}]M_S M_T}$$

$$= \sum_{m_{ij}}' \left\langle \begin{matrix} m_1 & m_2 & m_3 & m_4 \\ & m_{13} & m_{23} & m_{33} \\ & & m_{12} & m_{22} \\ & & & m_{11} \end{matrix} \middle| C^{(A\,B\,C)} \middle| \begin{matrix} m_1 & m_2 & m_3 & m_4 \\ & m_{13} & m_{23} & m_{33} \\ & & m_{12} & m_{22} \\ & & & m_{11} \end{matrix} \right\rangle$$

$$\qquad (6.12)$$

Here the prime on the summation symbol means as before that the values of m_{ij}, $1 \leq i$, $j \leq 3$, are restricted to those satisfying Eqs. (4.4a) and (4.4b). We then get the averages we are looking for in a recursive way,

$$\langle C^{(A\,B\,C)} \rangle^{[\tilde{m}]ST}$$
$$= [d([\mathbf{m}]ST)]^{-1} \Bigg\{ \langle\!\langle C^{(A\,B\,C)} \rangle\!\rangle^{[\tilde{m}]M_S=S\ M_T=T}$$
$$- \sum_{\substack{S' \geq S;\ T' \geq T \\ (S'T') \neq (ST)}} d([\mathbf{m}]S'T') \langle C^{(A\,B\,C)} \rangle^{[\tilde{m}]S'T'} \Bigg\}, \qquad (6.13)$$

starting from their value for the highest weight state

$$\langle C^{(A\,B\,C)} \rangle^{[\tilde{m}]PP'} = \langle\!\langle C^{(A\,B\,C)} \rangle\!\rangle^{[\tilde{m}]M_S=P\ M_T=P'}. \qquad (6.14)$$

Tables V and VI, VII and VIII, IX and X contain the averages of $C^{(111)}$, $C^{(202)}$, and $C^{(112)}$ respectively in all IR's appearing in s–d shell nuclei. In constructing these tables we have used extensively the symmetry relations satisfied by the averages in order to save space. We now proceed to derive them.

Symmetry relations of the first type relate averages calculated in the same U(4) IR but for different values of S and T. From the definition of the scalars $C^{(A\,B\,C)}$, we get immediately that

$$\langle C^{(111)} \rangle^{[\tilde{m}]S=a\ T=b} = \langle C^{(111)} \rangle^{[\tilde{m}]S=b\ T=a}, \qquad (6.15a)$$

$$\langle C^{(202)} \rangle^{[\tilde{m}]S=a\ T=b} = \langle C^{(022)} \rangle^{[\tilde{m}]S=b\ T=a}, \qquad (6.15b)$$

and

$$\langle C^{(112)} \rangle^{[\tilde{m}]S=a\ T=b} = \langle C^{(112)} \rangle^{[\tilde{m}]S=b\ T=a}. \qquad (6.15c)$$

In particular, Eq. (6.15b) shows that all the averages of $C^{(022)}$ can be deduced from those of $C^{(202)}$.

Symmetry relations of the second type relate averages calculated in different U(4) IR's but with the same values of S and T. First of all as the scalars $C^{(A\,B\,C)}$ belong to the enveloping algebra of $SU(4)$, their averages are the same in all U(4) IR's which are equivalent under SU(4), i.e.,

$$\langle C^{(A\,B\,C)} \rangle^{[\tilde{m}_1 \tilde{m}_2 \tilde{m}_3 \tilde{m}_4]ST} = \langle C^{(A\,B\,C)} \rangle^{[\widetilde{m_1-m_4}\ \widetilde{m_2-m_4}\ \widetilde{m_3-m_4}\ 0]ST}.$$
$$\qquad (6.16)$$

Secondly for any pair of contragredient IR's, we get from Eqs. (4.5) and (4.6) that

$$\langle C^{(111)} \rangle^{[\tilde{m}_1 \tilde{m}_2 \tilde{m}_3 \tilde{m}_4]ST} = - \langle C^{(111)} \rangle^{[\widetilde{m_1-m_4}\ \widetilde{m_1-m_3}\ \widetilde{m_1-m_2}\ 0]ST}, \qquad (6.17a)$$

and

$$\langle C^{(A\,B\,C)} \rangle^{[\tilde{m}_1 \tilde{m}_2 \tilde{m}_3 \tilde{m}_4]ST} = \langle C^{(A\,B\,C)} \rangle^{[\widetilde{m_1-m_4}\ \widetilde{m_1-m_3}\ \widetilde{m_1-m_2}\ 0]ST},$$
$$(A\,B\,C) = (202), \ (022), \ (112). \qquad (6.17b)$$

As a consequence of Eq. (6.17a), the averages of $C^{(111)}$ in any self-contragredient U(4) IR are equal to zero:

$$\langle C^{(111)} \rangle^{[\tilde{a}\ \tilde{b}\ \tilde{b}\ 0]ST} = 0. \qquad (6.18)$$

Let us quote finally some zero values of the averages of $C^{(A\,B\,C)}$ which result directly from the definition of these operators,

$$\langle C^{(111)} \rangle^{[\tilde{m}]S=0\ T} = \langle C^{(111)} \rangle^{[\tilde{m}]S\ T=0} = 0,$$

$$\langle C^{(202)} \rangle^{[\tilde{m}]S=0\ T} = \langle C^{(022)} \rangle^{[\tilde{m}]S\ T=0} = 0, \qquad (6.19)$$

and

$$\langle C^{(112)} \rangle^{[\tilde{m}]S=0\ T} = \langle C^{(112)} \rangle^{[\tilde{m}]S\ T=0} = 0.$$

By taking into account Eqs. (6.11), (6.15)–(6.19), and Tables V–X, the 28 systems of equations (6.10) can be explicitly written and solved for the $a_{\rho' r}$. The solution is given in Table XI. Introducing it into Eq. (6.8), we finally get explicit expressions for the operators $Q^u(\rho')$, which in conjunction with Tables V–X completely solve the problem of propagating the widths.

*Maître de recherches F.N.R.S.

[1] J. P. Elliott, Proc. Roy. Soc. A 245, 128, 562 (1958).

[2] V. Bargmann and M. Moshinsky, Nucl. Phys. 18, 967 (1960); 23, 177 (1961).

[3] M. Moshinsky and V. Syamala Devi, J. Math. Phys. 10, 455 (1969).

[4] R. T. Sharp and C. S. Lam, J. Math. Phys. 10, 2033 (1969).

[5] G. Racah, Lecture notes, "Group Theory and Spectroscopy," Institute for Advanced Study, Princeton, N.J., 1951; reprinted in Ergeb. Exact. Naturw. 37, 28 (1965).

[6] G. Racah, in Group Theoretical Concepts and Methods in Elementary Particles Physics, edited by F. Gürsey (Gordon and Breach, New York, 1964).

[7] B. R. Judd, W. Miller, Jr., J. Patera, and P. Winternitz, J. Math. Phys. 15, 1787 (1974).

[8] R. T. Sharp, J. Math. Phys. 16, 2050 (1975).

[9] J. B. French and K. F. Ratcliff, Phys. Rev. C 3, 94 (1971).

[10] C. Quesne, J. Math. Phys. 16, 2427 (1975).

[11]E. P. Wigner, Phys. Rev. **51**, 106 (1937).

[12]M. Moshinsky and J. G. Nagel, Phys. Letters **5**, 173 (1963).

[13]K. T. Hecht and S. C. Pang, J. Math. Phys. **10**, 1571 (1969).

[14]M. Brunet and M. Resnikoff, J. Math. Phys. **11**, 1474 (1970).

[15]J. P. Draayer, J. Math. Phys. **11**, 3225 (1970).

[16]R. U. Haq and J. C. Parikh, Nucl. Phys. A **220**, 349 (1974).

[17]M. Moshinsky, *Group Theory and the Many Body Problem* (Gordon and Breach, New York, 1968).

[18]N. Jacobson, *Lie Algebras* (Wiley, New York, 1962).

[19]H. Weyl, *The Classical Groups* (Princeton U. P., Princeton, N. J., 1946), pp. 251—52, 274—75.

[20]M. Hamermesh, *Group Theory* (Addison-Wesley, Reading, Mass., 1962).

[21]In Ref. 8, Sharp points out that this generating function was derived previously by Miller[22] and checked by him using an alternative procedure.

[22]W. Miller, Jr., unpublished.

[23]I. M. Gel'fand and M. L. Tseitlin, Dokl. Akad. Nauk SSSR **71**, 825 (1950).

[24]J. G. Nagel and M. Moshinsky, Rev. Mex. Fis. **14**, 29 (1965).

[25]B. G. Wybourne, *Symmetry Principles and Atomic Spectroscopy* (Wiley, New York, 1970).

Canonical realizations of the Poincaré group. II. Space–time description of two particles interacting at a distance, Newtonian-like equations of motion and approximately relativistic Lagrangian formulation

M. Pauri

*Istituto di Fisica dell'Università, Parma, Italy
and Istituto Nazionale di Fisica Nucleare, Sezione di Milano, Milano, Italy*

G. M. Prosperi

*Istituto di Fisica dell'Università, Milano, Italy
and Istituto Nazionale di Fisica Nucleare, Sezione di Milano, Milano, Italy*
(Received 11 March 1975)

The physical meaning of the relativistic action-at-a-distance dynamics for two particles in a canonical framework is investigated on the basis of a general formalism introduced in previous works. Starting from the well-known prescription given by Bakamjian and Thomas in terms of "center-of-mass" (**Q,P**) and "internal" (**ρ,π**) canonical coordinates, we show how to construct *physical*, i.e., covariant, position vectors $\mathbf{x}_r$ (Q, P, ρ, π) ($r = 1,2$) which approach the free particle coordinates in the limit $\rho \to \infty$ for short range forces; this procedure is actually performed by means of a $1/c^2$ power expansion for any interaction potential $U(\mathbf{\rho},\mathbf{\pi})$. In force of the zero-interaction theorem the physical coordinates, which do satisfy the world-line condition to any order in $1/c^2$, cannot play the role of canonical variables, i.e., the "localizability," $\{x_{ri}, x_{rj}\} = 0$ ($r = 1,2$), and the "causality" conditions $\{x_{ri}, x_{r'j}\} = 0$ ($\tau,\tau' = 1,2$; $\tau \neq \tau'$) cannot be simultaneously satisfied. It is possible, however, to satisfy the former set of equations to any order in $1/c^2$ by exploiting the arbitrariness lying in the definition of $\mathbf{x}_1$ and $\mathbf{x}_2$. By means of a suitable choice of a "gauge" for the internal variables, the remaining freedom is then shown to consist of the appearance of a single scalar function $\Lambda(\mathbf{\rho},\mathbf{\pi})$. This function, entering the defining relations of $\mathbf{x}_1$, $\mathbf{x}_2$ in terms of the canonical variables Q, P, ρ, π, plays the role of an additional interaction potential which is effective for the space–time description of the particles in the interaction region, but does not affect the scattering properties of the system. On the other hand, assuming a static nonrelativistic limit of the canonical potential, $U^{(0)} = U^{(0)}(\rho)$, the "causality" conditions are necessarily violated at the order of the radiation effects ($1/c^4$). In terms of $\mathbf{x}_1$, $\mathbf{x}_2$, the equations of motion assume a Newtonian-like structure $m_r \ddot{\mathbf{x}}_r = \mathbf{F}_r[\mathbf{x}_1 - \mathbf{x}_2, \mathbf{v}_1, \mathbf{v}_2]$ ($r = 1,2$), of the Currie type or a variety of manifestly covariant forms $m_r d^2 x_r^\mu / ds_r^2 = S^\mu{}_\nu f^\nu [x_1(s_1), x_2(s_2), u_1(s_1), u_2(s_2)]$, where $S^\mu{}_\nu$ is the Lorentz transformation which connects the laboratory frame with the Lorentz frame in which $x_1(s_1)$ and $x_2(s_2)$ are simultaneous. A final point is the derivation of the Newtonian-like equations of motion from a true Lagrangian variational principle $\delta \int L [\mathbf{x}_1, \mathbf{x}_2, \mathbf{v}_1, \mathbf{v}_2] dt = 0$. It is shown in general that if $U^{(0)}(\rho) \equiv 0$, this can be done only up to the post-Newtonian approximation, essentially because of the violation of the "causality" conditions at the order $1/c^4$. Then a general form of approximately relativistic Lagrangian for two particles is derived which actually contains all the examples quoted in the literature, among which the well-known Darwin–Breit and the Einstein–Infeld–Hoffmann Lagrangians. This investigation appears to disprove the widespread opinion according to which the zero-interaction theorem prevents the existence of invariant world lines and/or renders the relativity principle vacuous within a Hamiltonian framework.

1. INTRODUCTION

In a preceding paper[1] we have classified the canonical realizations[2–4] of the full Poincaré group and explicitly constructed various examples of physical interest. In particular, we have considered the nonirreducible realization corresponding to a system of two scalar particles and we have asserted that within our formalism one is led in a natural way to introduce a direct interaction between the particles according to the well-known prescription given by Bakamjian and Thomas.[5] Having defined suitable "center-of-mass" and "internal" variables Q, P and ρ, π, respectively, this prescription consists in assuming for the center-of-mass energy an expression of the form

$$Mc^2 \equiv c\pi_{10} + c\pi_{20} + U$$
$$= c\sqrt{m_1^2 c^2 + \pi^2} + c\sqrt{m_2^2 c^2 + \pi^2} + U(\rho, \pi), \qquad (1.1)$$

where the "potential" $U(\rho, \pi)$ is any rotationally in-

variant function of the internal variables. In the present paper we shall make the further assumption that $U_{[\kappa_1, \kappa_2]}(\rho, \pi)$ vanishes asymptotically for $|\rho| \to \infty$ and possibly depends in a *symmetrical way* on κ_1 and κ_2, which represent two sets of physical parameters characterizing the particles (masses, charges, and so on).

The theory based on the prescription (1.1) has received much attention in the literature (see in particular Refs. 6–8, 30, and the bibliography there contained) due to its simplicity and to the ability (already contained in Newtonian mechanics) that in principle has to treat problems of "microscopic" mechanics without the need for any detailed microscopic picture of the interaction. It is hardly necessary moreover to emphasize the importance of having a Hamiltonian (i. e., single time) description of relativistic particle dynamics. Not to speak of the calculation of constants

of motion, a Hamiltonian theory is relevant for the formulation of a relativistic statistical mechanics[7] and above all for the problem of quantization. The theory based on the Bakamjian–Thomas prescription appears to be the most general formulation of an instantaneous direct interaction relativistic mechanics consistent with a covariant formulation of the scattering problem[8]; finally Foldy[6] has shown that it is the most general solution of the Poincaré commutation relations for a system of particles with finite rest masses (and finite spin) for which an expansion in a series in $1/c^2$ exists.

In spite of these formal appealing features, however, a number of serious drawbacks of the theory have been put in evidence since its appearance and its physical relevance became substantially obscured. The difficulties are mainly connected with the problem of separability of the interaction[6,35] and with the consequences of the well-known zero-interaction theorem extablished in 1963 by Currie, Jordan, and Sudarshan.[9,35] The essential point is that the so-called internal canonical variables ρ, π cannot be truly *internal* as in the Galilei case and, if a direct interaction is present, they do not appear to have a clear physical meaning. As consequence there seems to be no natural basis in the theory for the proper relativistic space–time description of the interacting system. More precisely the situation is as follows. In the free particles case, $U \equiv 0$, the covariant position vectors q_1, q_2 of the particles and their linear momenta p_1, p_2 play the role of canonically conjugate variables. On the other hand, in the case $U \neq 0$, a system of canonical varaibles q_1', q_2', p_1', p_2' such that q_1' and q_2' transform correctly as covariant position vectors under the action of the Poincaré group cannot exist (zero-interaction theorem). As a consequence it has been argued that the theory is incompatible with the existence of invariant world lines.[5,10,6,7] At the same time it is believed that if the variables used to describe the configuration of the particles do not coincide with the basic canonical variables they are necessarily useless from a physical point of view and the relativity principle becomes vacuous.[8,11,35]

It will be shown in this paper that these difficulties are more apparent than real and that the theory can be given a definite physical content. Precisely we shall show that it is always possible to construct dynamical variables $x_1(Q, P, \rho, \pi)$, $x_2(Q, P, \rho, \pi)$ which at least in the case of potentials $U(\rho, \pi)$ which vanish fast enough in the limit $|\rho| \to \infty$, coincide asymptotically with the free particle position vectors $q_1(Q, P, \rho, \pi)$, $q_2(Q, P, \rho, \pi)$, respectively, and possessing all the relevant relativistic transformation properties *do define invariant world lines in the interaction region*. Clearly, the price to be paid for this result is that x_1, x_2 cannot represent the configurational part of a system of canonical variables.[12] In other words, the equations

$$\{x_{1i}, x_{1j}\} = 0, \quad \{x_{2i}, x_{2j}\} = 0, \tag{1.2}$$

$$\{x_{1i}, x_{2j}\} = 0, \quad i, j = 1, 2, 3, \tag{1.3}$$

cannot be all simultaneously satisfied. [See Bel[13] and Kunzle[13] for interesting alternatives.] It will appear that, far from preventing one from giving a physical foundation of the theory, this circumstance is strictly con-

nected with typical features of an action at a distance relativistic particle theory in the Hamiltonian form and essentially reflects the *hereditary character*[14] of the dynamics. At the same time the construction of the *physical*, i.e., *covariant* position variables within the Bakamjian–Thomas scheme will lead to very interesting connections of it with alternative formulations of the relativistic particle theory throwing new light on the inner mechanism of the zero-interaction theorem. [For example, it is not clear from the literature if the zero-interaction theorem does make any distinction between Eqs. (1.2) and (1.3).] Equations (1.2) and (1.3) will be called "localizability" and "causality" conditions, respectively.

After showing the *naturality* of the Bajamjian–Thomas prescription within our formalism (Sec. 2), the first problem dealt with in the paper (Sec. 3) is just the construction of the physical variables x_1, x_2 starting from the covariance or world-line condition[1,9]

$$\{K_i, x_{\tau j}\} = (1/c^2) x_{\tau i} \{x_{\tau j}, H\},$$

$$i, j = 1, 2, 3; \quad \tau = 1, 2. \tag{1.4}$$

Putting

$$x_\tau = Q + \xi_\tau(P, \rho, \pi), \quad \tau = 1, 2, \tag{1.5}$$

and replacing for H the expression $H = c(M^2 c^2 + P^2)^{1/2}$ with Mc^2 given by Eq. (1.1), the world line condition becomes a quasilinear first order partial differential equation for the vectors ξ_τ whose solutions are determined by Cauchy conditions of the form

$$\xi_\tau(o, \rho, \pi) \equiv \xi_{0\tau}(\rho, \pi), \quad \tau = 1, 2. \tag{1.6}$$

Then, assuming that the interaction potential can be expanded in a power series in the inverse squared light velocity

$$U_{[\kappa_1, \kappa_2]}(\rho, \pi) = U_{[\kappa_1, \kappa_2]}^{(0)}(\rho, \pi) + (1/c^2) U_{[\kappa_1, \kappa_2]}^{(1)}(\rho, \pi) + \cdots, \tag{1.7}$$

the covariant position vectors x_1, x_2 are explicitly constructed in terms of an expansion of the same kind (Sec. 4),

$$x_\tau = x_\tau^{(0)}(Q, P, \rho, \pi)$$
$$+ (1/c^2) x_\tau^{(1)}(P, \rho, \pi) + \cdots, \quad \tau = 1, 2. \tag{1.8}$$

At this stage the vectors $\xi_{0\tau}$'s are clearly arbitrary expressions of their arguments, apart from very general requirements. We show, however (Sec. 5), that *the vectors $\xi_{0\tau}(\rho, \pi)$, $\tau = 1, 2$, can always be chosen in order that the "localizability" conditions* (1.2) *be satisfied at any order of the $1/c^2$ expansion.* This determines the $\xi_{0\tau}(\rho, \pi)$ order by order up to the π-gradient of two scalar functions of the internal variables

$$\xi_{01}^{(n)}(\rho, \pi) = \xi_{01}^{*(n)}(\rho, \pi) + \frac{\partial \lambda_1^{(n)}(\rho, \pi)}{\partial \pi},$$

$$\xi_{02}^{(n)}(\rho, \pi) = \xi_{02}^{*(n)}(\rho, \pi) + \frac{\partial \lambda_2^{(n)}(\rho, \pi)}{\partial \pi}. \tag{1.9}$$

Actually a more significant role is played by the following functions simply related to those defined in (1.9), i.e.,

$$\lambda^{(n)} = \lambda_1^{(n)} - \lambda_2^{(n)},$$

$$\Lambda^{(n)} = (m_1 \lambda_1^{(n)} + m_2 \lambda_2^{(n)})/m, \tag{1.10}$$

where any $\Lambda^{(n)}_{[\kappa_1, \kappa_2]}(\rho, \pi)$ must be *antisymmetrical* in κ_1, κ_2. Precisely we show that the "relative" functions $\lambda^{(n)}(\rho, \pi)$ possess a "gauge" character in the sense that *they can always be disposed of order by order through a suitable canonical redefinition of the internal variables ρ, π and a renormalization of the interaction potential $U(\rho, \pi)$*, a procedure which can also be viewed as a glo global canonical equivalence on the whole realization of the group. On the other hand no restriction on the functional structure of the expressions $\Lambda^{(1)}(\rho, \pi)$, $\Lambda^{(2)}(\rho, \pi)$, $\cdots$ follows from the "localizability" conditions. Therefore we ask if there is any situation in which also the "causality" conditions could be verified to some extent by means of a suitable choice of the functions $\Lambda^{(1)}(\rho, \pi)$, $\Lambda^{(2)}(\rho, \pi)$, $\cdots$. First of all it is obvious that Eqs. (1.3) can be verified at zeroth order, i.e., in the nonrelativistic limit, for any interaction potential $U(\rho, \pi)$. At the first order in $1/c^2$ (Post-Newtonian approximation: PN) we find that they can be verified only under the assumption of a *static nonrelativistic limit* of the interaction, i.e., $U^{(0)} = U^{(0)}(\rho)$ by choosing for $\Lambda^{(1)}(\rho, \pi)$ the structure $\Lambda^{(1)} = (1/2m)\Omega^{(1)}_{[\kappa_1, \kappa_2]}(\rho) \cdot (\rho \cdot \pi)$. At the order $1/c^4$, the "causality" conditions can be satisfied by a similar choice of $\Lambda^{(2)}(\rho, \pi)$ and with $\Lambda^{(1)} \equiv 0$, only if $U^{(0)} \equiv 0$ and $U^{(1)} = U^{(1)}(\rho)$. The procedure can be easily iterated and shows that the "causality" conditions can be satisfied at any order of the expansion if and only if the interaction potential $U_{[\kappa_1, \kappa_2]}(\rho, \pi)$ is identically zero, a fact that appears as a *new proof of the zero-interaction theorem*. The result of this analysis is that, apart from the choice of the "gauge" $\lambda = (1/c^2)\lambda^{(1)} + (1/c^4)\lambda^{(2)} + \cdots$, essentially *two-independent functions of the internal vectors ρ, π*, namely the *canonical potential* $U(\rho, \pi)$ and the function $\Lambda(\rho, \pi)$ $= (1/c^2)\Lambda^{(1)}(\rho, \pi) + (1/c^4)\Lambda^{(2)}(\rho, \pi) + \cdots$, enter the relations which define the physical position vectors $\mathbf{x}_1, \mathbf{x}_2$ in terms of the basic canonical variables. From the point of view of the space—time description of the interacting particles represented by the time evolution of the vectors $\mathbf{x}_1, \mathbf{x}_2$, the particular structure of $\Lambda_{[\kappa_1, \kappa_2]}(\rho, \pi)$ must therefore be considered as a part of the specific dynamical theory in the sense that $\Lambda_{[\kappa_1, \kappa_2]}(\rho, \pi)$ plays the role of an *additional interaction potential* which, due to its symmetry properties, will be called *asymmetry* or *distinguishability potential*. [In particular for *identical* particles $\Lambda_{[\kappa_1, \kappa_2]}(\rho, \pi) \equiv 0$.] From the context of the discussion it will also be clear that, while this asymmetry potential is relevant for the space—time description of the system in the regions where the interaction is important, it does not affect the scattering properties which are completely accounted for by the *canonical potential* $U_{[\kappa_1, \kappa_2]}(\rho, \pi)$. The same is true for the dependence of the internal energy on possible "action variables" for the case of bounded motions of the particles. In this way the original canonical scheme appears to play the role of an asymptotic description as to an S-matrix theory.

The second problem dealt with in the present paper is the derivation of the equations of motion for the physical position vectors $\mathbf{x}_1, \mathbf{x}_2$ in a Newtonian-like form

$$m_\tau \ddot{\mathbf{x}}_\tau = \mathbf{F}_\tau[\mathbf{x}_1 - \mathbf{x}_2, \dot{\mathbf{x}}_1, \dot{\mathbf{x}}_2], \quad \tau = 1, 2, \tag{1.11}$$

where the forces $\mathbf{F}_\tau$ are again explicitly given by means

of an expansion in powers of $1/c^2$. The equations obtained provide a general explicit example of relativistically invariant differential equations of motion in the sense studied by Currie.[13,20] It is next shown that Eqs. (1.11) can be rewritten in the manifestly covariant form

$$\frac{d^2 x_\tau^\mu}{ds_\tau^2} = S_{(\tau)}{}^\mu{}_\nu f_{(\tau)}^\nu[x_1(s_1), x_2(s_2), u_1(s_1), u_2(s_2)], \tag{1.12}$$

where s_1 and s_2 are the proper times which specify two arbitrary points of the world lines of the two particles sharing a mutual spacelike relation and $S_{(\tau)}{}^\mu{}_\nu$ is the Lorentz transformation (which can be different for different τ's) which connects the laboratory frame with the Lorentz frame in which $x_1(s_1)$ and $x_2(s_2)$ are simultaneous. Equations (1.12) provide *infinitely many possible representations* of the equations of motion which are dynamically equivalent and are fixed by the particular assignment of the spacelike relation between $x_1(s_1)$ and $x_2(s_2)$ or, what is the same, of the functions $S_{(\tau)}{}^\mu{}_\nu$. Particular examples of this kind of description appear to be the manifestly covariant equations of motion studied by Havas and Plebański.[15]

The results achieved in Sec. 4, 5, and 6 allow one to reach a deeper understanding of the status of the so-called "approximately relativistic Lagrangians" which appear to have been extensively used in the literature and whose most famous examples are the Darwin—Breit "Lagrangian"[16] for electrodynamics and the Einstein—Infeld—Hoffmann "Lagrangian" (E. I. H.) derived in the slow-motion approximation from general relativity[17] which are both of the order $1/c^2$. Equations of the order $1/c^2$ have been applied to few-electron atoms and to two- or three-body motions of celestial bodies. However many-body applications are also possible as well as applications in magnetism and in the study of relativistic effects in statistical mechanics of charged particles.[7,18] The second part of Sec. 6 is devoted to the discussion of a possible Lagrangian formulation connected with our general canonical scheme. The essential point is that while, as shown by Hill and Kerner,[19,20] it is always possible to derive equations of the form (1.11) from a variational principle, the fact that the physical position variables cannot play the role of canonical variables (zero-interaction theorem) prevents the derivation of these same equations in the exact form from a *true Lagrangian* $L(\mathbf{x}_1, \mathbf{x}_2, \dot{\mathbf{x}}_1, \dot{\mathbf{x}}_2)$. Therefore the possibility of constructing a Lagrangian from which equations of motion of the form (1.11) can be derived only up to a given order in powers of $1/c^2$ *is directly connected to the possibility that the "localizability" and "causality" conditions* (1.2), (1.3) *are satisfied to the same order*. In particular *if we require having a nonvanishing interaction also in the nonrelativistic limit we can construct "Lagrangians" which reproduce the equations of motion only up to the PN approximation and furthermore only if the nonrelativistic limit interaction is static*. All the approximately relativistic Lagrangians discussed in the literature are just subjected to these limitations. It appears, however, that they have not been clearly stated (see for instance Ref. 21). (Curiously enough it is just this kind of unawareness that prevented an earlier dis-

covery of the zero-interaction theorem since the works of Darwin and Breit.)

An exhaustive investigation of the approximately relativistic Lagrangians derivable from a generalized Fokker principle has been recently given for a system of N particles by Woodcock and Havas[22] (W—H) (see also Ref. 23) who also discuss the interesting problem of the conditions under which an associated "adjunct field" theory (e. g., like the Feynman—Wheeler electrodynamics) exists. In order to make a comparison between our formulation and the W—H results it is convenient to restrict the form of the canonical potential assuming the structure

$$U = \gamma \hat{U}_{[\kappa_1, \kappa_2]} \left[\frac{\gamma}{\mu c^2}, \frac{\rho}{l}, \frac{\pi^2}{\mu^2 c^2}, \frac{\sigma^2}{l^2 \mu^2 c^2} \right]$$

$$\equiv U^{(0)}_{[\kappa_1, \kappa_2]}(\rho) + \frac{1}{c^2} \left\{ \Xi_{[\kappa_1, \kappa_2]}(\rho) + \frac{\pi^2}{2\mu^2} \Phi_{[\kappa_1, \kappa_2]}(\rho) \right.$$

$$\left. + \frac{\sigma^2}{2\mu^2} \Psi_{[\kappa_1, \kappa_2]}(\rho) \right\} + O\left(\frac{1}{c^4}\right), \quad \mu = \frac{m_1 m_2}{m}, \quad \sigma = \rho \cdot \pi,$$

(1.13)

where $\hat{U}_{[\kappa_1, \kappa_2]}$ is a dimensionless function which contains the minimum possible number of dimensional constants besides the parameters κ_1, κ_2, namely γ and l with the dimensions of an energy and a length, respectively. On this basis we derive a very general approximately relativistic Lagrangian [Eqs. (6.39)—(6.41)] in which four generalized interaction potentials $\Omega^{(1)}_{[\kappa_1, \kappa_2]}(r)$, $\tilde{\Phi}_{[\kappa_1, \kappa_2]}(r)$, $\tilde{\Psi}_{[\kappa_1, \kappa_2]}(r)$, $\tilde{\Xi}_{[\kappa_1, \kappa_2]}(r)$, $(r = |\mathbf{x}_1 - \mathbf{x}_2|)$ occur, in addition to the static Newtonian interaction $U^{(0)}_{[\kappa_1, \kappa_2]}(r)$. The "distinguishability" potential $\Omega^{(1)}_{[\kappa_1, \kappa_2]}(r)$ appears to have a particular interest in this context. This potential follows directly from the validity of the "causality" conditions at the PN approximation [which impose $\Lambda^{(1)} = (1/2m)\Omega^{(1)}_{[\kappa_1, \kappa_2]}(r) \cdot \sigma$] and is antisymmetrical in the particle parameters κ_1, κ_2. The appearance of $\Omega^{(1)}_{[\kappa_1, \kappa_2]}(r)$ shows in turn the dynamical role of Eqs. (1.3). Our approximately relativistic Lagrangians essentially coincide with the class obtained by W—H if we set

$$\tilde{\Xi}(r) = \frac{1}{m} U^{(0)}(r) \frac{1}{r} \frac{dU^{(0)}}{dr}$$

(1.14)

and make suitable identifications for the potentials $U^{(0)}_{[\kappa_1, \kappa_2]}(r)$, $\tilde{\Phi}_{[\kappa_1, \kappa_2]}(r)$, $\tilde{\Psi}_{[\kappa_1, \kappa_2]}(r)$, $\Omega^{(1)}_{[\kappa_1, \kappa_2]}(r)$. On the other hand it is interesting to find that corresponding to suitable, more general, choices of the potential $\tilde{\Xi}_{[\kappa_1, \kappa_2]}(r)$, our scheme actually contains just all the approximately relativistic Lagrangians for two particles known in the literature, i.e., in particular also "Lagrangians" corresponding to field theories which are typically nonlinear at the PN approximation (and thus cannot belong to the W—H classification; see however P. Havas and J. Stachel, Ref. 23) such as the already mentioned E. I. H. gravitational "Lagrangian" and that given by Bazański[24] which describes the general relativistic theory of two gravitating charged particles. Let us remark moreover that, in force of our Hamiltonian formulation in terms of the basic canonical variables $\mathbf{Q}, \mathbf{P}, \rho, \pi$, the approximate constants of motion which are characteristic of the approximately relativistic Lagrangian formulations[23, 25] result in just the PN approximations of the exact constants of motion of our

equations and the approximately relativistic Lagrangians themselves correspond to PN approximations of exact variational principles.

As a final consideration, we recall that Hill and Kerner[19, 20] have shown that equations of motion of Newtonian-like form can always be put in Hamiltonian form, while Kerner[26] (see also Ref. 20) has shown that the integro-differential equations of motions associated to a variational principle of the Fokker type can always be reduced to a Newtonian-like structure provided their solutions are assumed to admit an expansion in powers of $1/c^2$ and possess some kind of analytical contiguity to the free motions. Therefore our analysis establishes in some way the essential equivalence and the definite relations existing among the various approaches discussed in the literature. (It should be worth stressing that the convergence of the formal expansion in powers of $1/c^2$ is an open question despite the very popular use made of it in the literature. Its naturalness, of course, is based on the fact that in the limit $c \to \infty$ it is sufficient to effect the transition from the Poincaré group to the Galilei group.) Precisely:

(a) the instant-form canonical approach initiated by Dirac and Bakamjian—Thomas[5, 6];

(b) the Newtonian-like differential approach introduced by Currie[14];

(c) the manifestly covariant equations approach discussed by Havas—Plebański[15];

(d) the variational approach based on generalized Fokker principles which was first emphasized by the Feynman—Wheeler electrodynamics and successively generalized by Van Dam—Wigner and others.[10, 19, 20, 22, 26, 27, 38]

Our approach allows in principle for a canonization of these theories.

2. INTERACTION AT A DISTANCE BETWEEN TWO SCALAR PARTICLES IN A HAMILTONIAN FRAMEWORK

The canonical generators of a realization of the Poincaré group corresponding to a system of two free scalar particles can be written

$$\mathbf{J} = \mathbf{q}_1 \wedge \mathbf{p}_1 + \mathbf{q}_2 \wedge \mathbf{p}_2, \quad \mathbf{K} = -\frac{p_{10}}{c} \mathbf{q}_1 - \frac{p_{20}}{c} \mathbf{q}_2,$$

$$\mathbf{T} = \mathbf{p}_1 + \mathbf{p}_2,$$

(2.1)

$$H = c(p_{10} + p_{20}), \quad p_{i0} = \sqrt{m_i^2 c^2 + \mathbf{p}_i^2}, \quad i = 1, 2;$$

where the variables $\mathbf{q}_1, \mathbf{p}_1, \mathbf{q}_2, \mathbf{p}_2$ have an obvious physical meaning (see Ref. 1). Introducing the "center-of-mass" variables

$$\mathbf{P} = \mathbf{p}_1 + \mathbf{p}_2,$$

$$\mathbf{Q} = \frac{p_{10}(Mc^2 + H) - c\mathbf{p}_1 \cdot \mathbf{P}}{Mc(Mc^2 + H)} \mathbf{q}_1 + \frac{p_{20}(Mc^2 + H) - c\mathbf{p}_2 \cdot \mathbf{P}}{Mc(Mc^2 + H)} \mathbf{q}_2$$

$$+ \frac{c(\mathbf{q}_1 - \mathbf{q}_2) \cdot \mathbf{P}}{MH(Mc^2 + H)} (p_{20}\mathbf{p}_1 - p_{10}\mathbf{p}_2),$$

(2.2)

and the "internal" variables

$$\pi = \frac{p_{20}(Mc^2 + H) - c\mathbf{p}_2 \cdot \mathbf{P}}{Mc(Mc^2 + H)} \mathbf{p}_1 - \frac{p_{10}(Mc^2 + H) - c\mathbf{p}_1 \cdot \mathbf{P}}{Mc(Mc^2 + H)} \mathbf{p}_2$$

1471   J. Math. Phys., Vol. 17, No. 8, August 1976

M. Pauri and G.M. Prosperi   1471

$$\rho = q_1 - q_2 + (q_1 - q_2) \cdot \mathbf{P} \qquad (2.3)$$

$$\times \left[\frac{\mathbf{P}}{M(Mc^2 + H)} + \frac{1}{Mc}\left(\frac{1}{\pi_{10}} - \frac{1}{\pi_{20}} - \frac{c\pi \cdot \mathbf{P}}{H\pi_{10}\pi_{20}}\right)\pi \right],$$

where

$$M = (1/c^2)\sqrt{H^2 - c^2\mathbf{P}^2} \quad \text{and} \quad \pi_{i0} = \sqrt{m_i^2 c^2 + \pi^2}, \quad i = 1, 2,$$

Eqs. (2.1) become [see Ref. 1, Eqs. (4.8), (4.9)]

$$\mathbf{J} = \mathbf{Q} \wedge \mathbf{P} + \mathbf{S}, \quad \mathbf{K} = -\left(\frac{H}{c^2}\right)\mathbf{Q} + \frac{\mathbf{S} \wedge \mathbf{P}}{Mc^2 + H},$$

$$\mathbf{T} = \mathbf{P}, \quad H = c\sqrt{M^2 c^2 + \mathbf{P}^2}, \qquad (2.4)$$

where

$$\mathbf{S} = \rho \wedge \pi, \qquad (2.5)$$

$$Mc^2 = c(\pi_{10} + \pi_{20})$$
$$= c\sqrt{m_1^2 c^2 + \pi^2} + c\sqrt{m_2^2 c^2 + \pi^2}. \qquad (2.6)$$

The physical meaning of the variables $\mathbf{Q}, \mathbf{P}, \rho, \pi$ has been discussed in detail in Ref. 1.

According to Bakamjian and Thomas[5] an interaction can be introduced between the particles by simply maintaining the formal structure of Eqs. (2.4), (2.5) and replacing Eq. (2.6) by

$$Mc^2 = c\sqrt{m_1^2 c^2 + \pi^2} + c\sqrt{m_2^2 c^2 + \pi^2} + U(\rho, \pi), \qquad (2.7)$$

where $U(\rho, \pi)$ is a function which plays here the role of the nonrelativistic potential and which is assumed to be rotationally invariant,

$$\{J_i, U(\rho, \pi)\} = \{S_i, U(\rho, \pi)\} = 0, \qquad (2.8)$$

i.e., to be a function of $\rho = |\rho|$, $\pi = |\pi|$, $\sigma \equiv \rho \cdot \pi$ only.

This procedure appears to be quite natural within the framework of the formalism of the canonical realizations developed in Refs. 1—4 and it turns out to be essentially unique under some reasonable hypotheses.

We assume first that the state of the system can be always characterized by a set of twelve canonical variables like $q_1, \mathbf{p}_1, q_2, \mathbf{p}_2$ or $\mathbf{Q}, \mathbf{P}, \rho, \pi$. Then the introduction of an interaction clearly amounts to changing the structure of the canonical generators as given by Eqs. (2.1) or (2.4)—(2.6) leaving invariant the *transformation properties* of the quantities which are *physically relevant*. With this in view the most natural restriction appears to assume that $q_1, \mathbf{p}_1, q_2, \mathbf{p}_2$ or $\mathbf{Q}, \mathbf{P}, \rho, \pi$ maintain their transformation properties under the Euclidean subgroup of the space rotations and translations

$$\{T_i, q_{1j}\} = \{T_i, q_{2j}\} = -\delta_{ij}$$
$$\{T_i, p_{1j}\} = \{T_i, p_{2j}\} = 0, \qquad (2.9)$$
$$\{J_i, q_{\tau j}\} = \epsilon_{ijk}q_{\tau k}$$
$$\{J_i, p_{\tau j}\} = \epsilon_{ijk}p_{\tau k}, \quad \tau = 1, 2,$$

$$\{T_i, Q_j\} = -\delta_{ij},$$
$$\{T_i, P_j\} = \{T_i, \rho_j\} = \{T_i, \pi_j\} = 0, \qquad (2.10)$$
$$\{J_i, Q_j\} = \epsilon_{ijk}Q_k, \quad \{J_i, P_j\} = \epsilon_{ijk}P_k,$$
$$\{J_i, \rho_j\} = \epsilon_{ijk}\rho_k, \quad \{J_i, \pi_j\} = \epsilon_{ijk}\pi_k.$$

On the other hand the zero-interaction theorem established by Currie, Jordan and Sudarshan[9] prevents the possibility of modifying in any nontrivial way the structure of the canonical generators leaving invariant the transformation properties of q_1 and q_2 also under the special Lorentz transformations, i.e., Eqs. (1.4) in addition to Eqs. (2.9). Consequently q_1 and q_2 can no longer represent the positions of the relativistic particles in the presence of a direct interaction and their physical meaning will not be simple in general. On the other hand variables like $\mathbf{Q}, \mathbf{P}, \rho, \pi$ satisfying Eqs. (2.10) appear more suitable to the present considerations. Actually it will be always possible to choose $\mathbf{Q}$ and $\mathbf{P}$ so that they coincide with the variables $\mathfrak{Q}$ and $\mathfrak{P}$ of the scheme A (see Ref. 1, Table A_1) even in the case of interaction, i.e.,

$$\mathfrak{Q} = \mathbf{Q}, \quad \mathfrak{P} = \mathbf{P}. \qquad (2.11)$$

This implies that Eqs. (2.4) remain unchanged and that $\mathbf{S}$ and Mc are functions of ρ, π only. Equations (2.10), last line, are then equivalent to

$$\{S_i, \rho_j\} = \epsilon_{ijk}\rho_k, \quad \{S_i, \pi_j\} = \epsilon_{ijk}\pi_k, \qquad (2.12)$$

which imply that $\mathbf{S}$ retains its standard form (2.5). Therefore only the quantity Mc as a function of the internal variables ρ, π can be modified. Finally, in order that the fundamental Poisson brackets among the canonical generators of the group are preserved, we must impose the condition

$$\{S_i, Mc(\rho, \pi)\} = 0, \qquad (2.13)$$

which limits Mc to be a function of ρ, π, σ only. This condition coincides with the Bakamjian—Thomas prescription if $U(\rho, \pi)$ is suitably defined. We shall require in addition that $\mathbf{Q}, \mathbf{P}, \rho, \pi$ transform under the discrete operations of the full Poincaré group as in the free case, i.e.,

$$I_s\mathbf{Q} = -\mathbf{Q}, \quad I_s\mathbf{P} = -\mathbf{P}, \quad \text{canonical space reflection} \quad (2.14)$$
$$I_s\rho = -\rho, \quad I_s\pi = -\pi$$

$$I_t^*\mathbf{Q} = \mathbf{Q}, \quad I_t^*\mathbf{P} = -\mathbf{P}, \quad \text{anticanonical time reflection} \quad (2.15)$$
$$I_t^*\rho = \rho, \quad I_t^*\pi = -\pi.$$

Then Eqs. (2.4), (2.5), (2.7) do admit a realization of the full group provided $U(\rho, \pi)$ is an even function of σ [see Ref. 1, Eqs. (2.17), and Sec. 3f].

Let us stress that the conditions (2.10) and (2.12) *characterize ρ and π up to a canonical transformation which preserves their vectorial nature and consequently is generated by a scalar function of the ρ, π themselves.* Since the interaction must become negligible when the particles are sufficiently far apart, it is natural to assume in force of the discussion made in Ref. 1 (Sec. 4a) on the physical meaning of ρ and π that these "internal" vectors can be chosen in such a way that ρ coincides asymptotically with the relative coordinate of the particles in the center-of-mass system and the expressions q_1, q_2, obtained from Eqs. (2.2), (2.3) by inversion, can be identified with the position coordinates of the particles in this same limit. These requirements are equivalent to assuming that the vectors

ρ, π can be chosen in such a way that the interaction potential $U(\rho, \pi)$ vanishes fast enough for $\rho \to \infty$,

$$U(\rho, \sigma, \pi) \underset{\rho \to \infty}{\to} 0. \tag{2.16}$$

This is a further hypothesis we shall generally make in the present work.

3. PHYSICAL POSITION VARIABLES FOR TWO PARTICLES INTERACTING AT A DISTANCE

As announced in the Introduction, we want to discuss the problem of the space—time description of the interacting particles in terms of physical, i.e., covariant, position vectors $\mathbf{x}_1, \mathbf{x}_2$:

$$\mathbf{x}_1 = \mathbf{x}_1(\mathbf{Q}, \mathbf{P}, \rho, \pi), \tag{3.1}$$
$$\mathbf{x}_2 = \mathbf{x}_2(\mathbf{Q}, \mathbf{P}, \rho, \pi).$$

We recall[1] that the transformation laws under space translations, space rotations, and special Lorentz transformations for a covariant position vector are expressed within a canonical realization of the Poincaré group by the following relations:

$$\{P_i, x_j\} = - \delta_{ij}, \tag{3.2}$$

$$\{J_i, x_j\} = \epsilon_{ijk} x_k, \tag{3.3}$$

$$\{K_i, x_j\} = \frac{1}{c^2} x_i \{x_j, H\}. \tag{3.4}$$

Similarly, the transformation properties under space and time reflections are expressed by

$$I_s \mathbf{x} = - \mathbf{x}, \tag{3.5}$$

$$I_t^* \mathbf{x} = \mathbf{x}. \tag{3.6}$$

Our problem will consist in finding appropriate solutions of Eqs. (3.2)— (3.6).

Equation (3.2) directly implies that

$$\mathbf{x} = \mathbf{Q} + \xi(\mathbf{P}, \rho, \pi). \tag{3.7}$$

Then Eq. (3.3) says that ξ must be a vector under space-rotations. Finally, introducing the expression (3.7) in Eq. (3.4), after a few manipulations we obtain the following quasilinear equation for the vector ξ:

$$\frac{\partial \xi_j}{\partial P_i} = - \frac{1}{H} \left(\xi_i - \frac{(\mathbf{S} \wedge \mathbf{P})_i}{M(Mc^2 + H)} \right) \{\xi_j, H\}$$
$$- \frac{c^2}{H^2} \xi_i P_j - \frac{c^2 \epsilon_{itk} P_i}{H(Mc^2 + H)} \{S_k, \xi_j\}$$
$$+ \frac{c^2 \epsilon_{ijk} S_k}{H(Mc^2 + H)} + \frac{c^4 (\mathbf{S} \wedge \mathbf{P})_i P_j}{H^2 (Mc^2 + H)^2}. \tag{3.8}$$

It can be checked that this equation is integrable. Then its solutions are determined by means of Cauchy conditions of the form[28]

$$\xi_i(o, \rho, \pi) = \xi_{0i}(\rho, \pi). \tag{3.9}$$

If we choose $\xi_{0i}(\rho, \pi) \equiv 0$, Eq. (3.8) furnishes the expression of the covariant center of mass $\mathbf{X}^1$ in terms of the basic canonical variables. On the other hand if we assume $U(\rho, \pi) \equiv 0$ (free particles case), and consider two different solutions ξ_1, ξ_2 of Eq. (3.8) corresponding to the Cauchy conditions

$$\xi_1(o, \rho, \pi) = [\pi_{20}/(\pi_{10} + \pi_{20})] \rho, \tag{3.10}$$
$$\xi_2(o, \rho, \pi) = [- \pi_{10}/(\pi_{10} + \pi_{20})] \rho,$$

we obtain just the free particle coordinates

$$\mathbf{x}_1 = \mathbf{q}_1(\mathbf{Q}, \mathbf{P}, \rho, \pi), \quad \mathbf{x}_2 = \mathbf{q}_2(\mathbf{Q}, \mathbf{P}, \rho, \pi), \tag{3.11}$$

where

$$\mathbf{q}_1 = \mathbf{Q} - \frac{\rho \cdot \mathbf{P}}{M(Mc^2 + H)} \frac{c\pi_{10}\pi_{20}}{\pi_{10}H + c\pi \cdot \mathbf{P}} \mathbf{P} + \frac{1}{Mc}$$
$$\times \left(\pi_{20} - \frac{c\pi \cdot \mathbf{P}}{Mc^2 + H} \right) \rho + \frac{\rho \cdot \mathbf{P}}{M} \left(\frac{1}{Mc^2 + H} - \frac{\pi_{20}}{\pi_{10}H + c\pi \cdot \mathbf{P}} \right) \pi, \tag{3.12}$$

$$\mathbf{q}_2 = \mathbf{Q} + \frac{\rho \cdot \mathbf{P}}{M(Mc^2 + H)} \frac{c\pi_{10}\pi_{20}}{\pi_{20}H - c\pi \cdot \mathbf{P}} \mathbf{P} - \frac{1}{Mc}$$
$$\times \left(\pi_{10} + \frac{c\pi \cdot \mathbf{P}}{Mc^2 + H} \right) \rho + \frac{\rho \cdot \mathbf{P}}{M} \left(\frac{1}{Mc^2 + H} - \frac{\pi_{10}}{\pi_{20}H - c\pi \cdot \mathbf{P}} \right) \pi,$$

which can also be obtained by inversion of Eqs. (2.2), (2.3) [see Ref. 1, Eqs. (4.7)].

In the presence of an interaction, we shall characterize the position vectors $\mathbf{x}_1, \mathbf{x}_2$ by means of two Cauchy conditions of the form

$$\xi_1(o, \rho, \pi) = \frac{\pi_{20}}{\pi_{10} + \pi_{20}} \rho + \chi_1(\rho, \pi),$$
$$\xi_2(o, \rho, \pi) = - \frac{\pi_{10}}{\pi_{10} + \pi_{20}} \rho + \chi_2(\rho, \pi). \tag{3.13}$$

The vectors χ_1 and χ_2 will be required to satisfy a number of general conditions:

(a) *reflection conditions*:

From Eqs. (3.5)—(3.7), (3.13), owing to Eqs. (2.14), (2.15), it follows that

$$\chi_\tau(\rho, \pi) = - \chi_\tau(- \rho, \pi) = \chi_\tau(\rho, - \pi), \quad \tau = 1, 2. \tag{3.14}$$

These relations imply in particular that the vectors χ_τ must have the form

$$\chi_\tau(\rho, \pi) = \alpha_\tau(\rho, \pi, \sigma)\rho + \beta_\tau(\rho, \pi, \sigma)\pi, \quad \tau = 1, 2, \tag{3.15}$$

where α_τ and β_τ are even and odd functions of σ, respectively;

(b) *symmetry conditions*:

We assume that the *individual properties* of the particles interacting at a distance can be essentially characterized in terms of a *finite number of parameters* (masses, charges and so on) which will be collectively denoted by the symbols κ_1 and κ_2, respectively; that the interaction potential $U_{[\kappa_1, \kappa_2]}(\rho, \pi)$ depends *symmetrically* on κ_1 and κ_2; and, consistently, that *the world lines of the two particles are left invariant under the exchange of the positions, velocities, and values of the physical parameters of the particles, made at a given time*. The last requirements implies that in a reference frame corresponding to $\mathbf{P} = o$, $\mathbf{Q} = o$ (e.g., at $t = 0$), the values of $\mathbf{x}_1$ and $\mathbf{x}_2$ must interchange under the operation

$$Z : \begin{cases} \rho \to - \rho, \\ \pi \to - \pi, \\ \kappa_1 \rightleftarrows \kappa_2, \end{cases} \tag{3.16}$$

i.e., a space reflection followed by an interchange of the values of the individual parameters. This implies in turn the exchange of χ_1 with χ_2, or, due to the reflection conditions (3.14), the *symmetry condition*

$$\chi_2(\rho,\pi,\kappa_1,\kappa_2) = -\chi_1(\rho,\pi,\kappa_2,\kappa_1); \qquad (3.17)$$

(c) *locally vanishing interaction and asymptotic conditions*:

The vectors $\chi_\tau(\rho,\pi)$, $\tau = 1,2$, are required to vanish locally in those regions of the phase-space where the interaction potential $U(\rho,\pi)$ itself does. In particular, according to the discussion made in Sec. 1 about the asymptotic condition for $\rho \to \infty$, we shall assume that, *for short-range interactions at least*

$$[\mathbf{x}_\tau(\mathbf{Q},\mathbf{P},\rho,\pi) - \mathbf{q}_\tau(\mathbf{Q},\mathbf{P},\rho,\pi)]_{\rho\to\infty} \to o, \quad \tau = 1,2, \qquad (3.18)$$

with the consequence that

$$\chi_\tau(\rho,\pi)_{\rho\to\infty} \to o, \quad \tau = 1,2; \qquad (3.19)$$

(d) *nonrelativistic limit*:

Finally, the usual nonrelativistic expressions for $\mathbf{x}_1$ and $\mathbf{x}_2$ in terms of the "center of mass" and "relative" coordinates should be recovered in the limit $c \to \infty$,

$$\begin{aligned}\mathbf{x}_{1_{n.r.}} &= \mathbf{Q}_{n.r.} + (m_2/m)\boldsymbol{\rho}_{n.r.},\\ \mathbf{x}_{2_{n.r.}} &= \mathbf{Q}_{n.r.} - (m_1/m)\boldsymbol{\rho}_{n.r.},\end{aligned} \quad m = m_1 + m_2. \qquad (3.20)$$

Therefore we shall require

$$\chi_\tau(\rho,\pi)_{c\to\infty} \to o, \quad \tau = 1,2. \qquad (3.21)$$

Once these general requirements have been taken into account, *the definite choice of the vectors* $\chi_1(\rho,\pi)$ *and* $\chi_2(\rho,\pi)$ *must be considered as a part of a specific dynamical theory* for the particle world lines. As a matter of fact, it is clear that, *for a given interaction potential* $U(\rho,\pi)$, *different choices of the* $\chi_\tau(\rho,\pi)$ *correspond to different functional expressions of the physical position variables* $\mathbf{x}_1, \mathbf{x}_2$ *in terms of the basic canonical variables* $\mathbf{Q}, \mathbf{P}, \rho, \pi$ *and therefore in general to different world lines of the particles*. Of course the most simple possibility for making the requirements (a)—(d) satisfied is to choose

$$\chi_1(\rho,\pi) = \chi_2(\rho,\pi) \equiv o. \qquad (3.22)$$

This choice has the simple consequence that the physical coordinates $\mathbf{x}_1, \mathbf{x}_2$ coincide with the free particle position variables $\mathbf{q}_1, \mathbf{q}_2$ in the center-of-mass system,

but it does not appear to have any clear physical motivation. On the other hand, we shall see in the following that a very far-reaching program is to exploit the formal arbitrariness of the $\chi_\tau(\rho,\pi)$ in order that the "localizability" and "causality" conditions (1.2) and (1.3) are possibly satisfied to some extent. This point is discussed in detail in Sec. 5.

It is convenient for later use to split the vectors $\chi_1(\rho,\pi)$, $\chi_2(\rho,\pi)$ into "center-of-mass" and "relative" parts. We put [see Eq. (3.7), (3.13)]

$$\begin{aligned}\chi_1 &\equiv X + [\pi_{20}/(\pi_{10}+\pi_{20})]\chi,\\ \chi_2 &\equiv X - [\pi_{10}/(\pi_{10}+\pi_{20})]\chi,\end{aligned} \qquad (3.23)$$

or

$$\chi \equiv \chi_1 - \chi_2, \quad X \equiv \frac{\pi_{10}\chi_1 + \pi_{20}\chi_2}{\pi_{10}+\pi_{20}}. \qquad (3.24)$$

In the center-of-mass system $(\mathbf{P} \equiv o)$ we have then

$$\begin{aligned}\mathbf{Q} &= \frac{\pi_{10}\mathbf{x}_{1_{c.m.}} + \pi_{20}\mathbf{x}_{2_{c.m.}}}{\pi_{10}+\pi_{20}} - X,\\ \rho &= \mathbf{x}_{1_{c.m.}} - \mathbf{x}_{2_{c.m.}} - \chi,\end{aligned} \qquad (3.25)$$

[where $\mathbf{x}_{\tau_{c.m.}}$ means $\mathbf{x}_\tau(\mathbf{Q},\mathbf{P},\rho,\pi)|_{\mathbf{P}\equiv o}$, $\tau = 1,2$] which clarifies the physical meaning of χ and X. In terms of these expressions, the symmetry conditions (b) [Eqs. (3.17)] takes the form

$$\begin{aligned}\chi(\rho,\pi;\kappa_1,\kappa_2) &= \chi(\rho,\pi;\kappa_2,\kappa_1),\\ X(\rho,\pi;\kappa_1,\kappa_2) &= -X(\rho,\pi;\kappa_2,\kappa_1),\end{aligned} \qquad (3.26)$$

with the consequence that *in the case of identical particles* $(\kappa_1 \equiv \kappa_2)$ we must have

$$X(\rho,\pi) \equiv o, \qquad (3.27)$$

i.e., in the center-of-mass system

$$\mathbf{Q} = (\mathbf{x}_{1_{c.m.}} + \mathbf{x}_{2_{c.m.}})/2. \qquad (3.28)$$

Note that, in the case of the simple choice $\chi_1(\rho,\pi) \equiv \chi_2(\rho,\pi) \equiv o$, Eq. (3.25) becomes

$$\begin{aligned}\mathbf{Q} &= \frac{\pi_{10}\mathbf{x}_{1_{c.m.}} + \pi_{20}\mathbf{x}_{2_{c.m.}}}{\pi_{10}+\pi_{20}},\\ \rho &= \mathbf{x}_{1_{c.m.}} - \mathbf{x}_{2_{c.m.}},\end{aligned} \qquad (2.34)$$

which expresses in particular the fact that the interparticle separation in the center-of-mass system maintains its canonical character also in the presence of the interaction.

4. EXPLICIT CONSTRUCTION OF THE PHYSICAL POSITION VARIABLES BY MEANS OF AN EXPANSION IN POWERS OF $1/c^2$

We assume now that the interaction potential $U_{[\kappa_1,\kappa_2]}(\rho,\pi)$ is represented in the form

$$U_{[\kappa_1,\kappa_2]}(\rho,\pi) = U^{(0)}_{[\kappa_1,\kappa_2]}(\rho,\pi) + (1/c^2)U^{(1)}_{[\kappa_1,\kappa_2]}(\rho,\pi) + \cdots \qquad (4.1)$$

and consider the problem of the actual construction of the vectors $\mathbf{x}_1, \mathbf{x}_2$. We put

$$\mathbf{x}_\tau = \mathbf{x}_\tau^{(0)} + (1/c^2)\mathbf{x}_\tau^{(1)} + (1/c^4)\mathbf{x}_\tau^{(2)} + \cdots = \mathbf{Q} + \xi_\tau^{(0)} + (1/c^2)\xi_\tau^{(1)} + (1/c^4)\xi_\tau^{(2)} + \cdots, \quad \tau = 1,2, \qquad (4.2)$$

and

$$\chi_\tau = (1/c^2)\chi_\tau^{(1)} + (1/c^4)\chi_\tau^{(2)} + \cdots, \quad \tau = 1,2, \qquad (4.3)$$

where the nonrelativistic limit [condition (d)] of Sec. 2 has already been enforced by choosing $\chi_\tau^{(0)} \equiv o$. Then we have

$$H = mc^2 + H^{(0)} + (1/c^2)H^{(1)} + (1/c^4)H^{(2)} + \cdots, \quad \mathbf{K} = \mathbf{K}^{(0)} + (1/c^2)\mathbf{K}^{(1)} + (1/c^4)\mathbf{K}^{(2)} + \cdots, \qquad (4.4)$$

where

$$H^{(0)} = \mathbf{P}^2/2m + \pi^2/2\mu + U^{(0)}(\rho, \pi),$$

$$H^{(1)} = -\left[\frac{1}{8}\frac{\mathbf{P}^4}{m^3} + \frac{\mathbf{P}^2}{2m^2}\left(\frac{\pi^2}{2\mu} + U^{(0)}(\rho, \pi)\right) + \frac{1}{8}\left(\frac{1}{m_1^3} + \frac{1}{m_2^3}\right)\pi^4\right] + U^{(1)}(\rho, \pi),$$

$$H^{(2)} = \left\{\frac{1}{16}\frac{\mathbf{P}^6}{m^5} + \frac{3\mathbf{P}^4}{8m^4}\left(\frac{\pi^2}{2\mu} + U^{(0)}(\rho, \pi)\right) + \frac{\mathbf{P}^2}{2m^2}\left[\frac{1}{8}\left(\frac{1}{m_1^3} + \frac{1}{m_2^3}\right) + \frac{1}{4m\mu^2}\right]\pi^4\right.$$

$$\left. + \frac{1}{16}\left(\frac{1}{m_1^5} + \frac{1}{m_2^5}\right)\pi^6\right\} - \frac{\mathbf{P}^2}{2m^2}U^{(1)}(\rho, \pi) + U^{(2)}(\rho, \pi), \ldots,$$ (4.5)

and

$$\mathbf{K}^{(0)} = -m\mathbf{Q}, \quad \mathbf{K}^{(1)} = -\left(\frac{\mathbf{P}^2}{2m} + \frac{\pi^2}{2\mu} + U^{(0)}(\rho, \pi)\right)\mathbf{Q} + \frac{\mathbf{S}\wedge\mathbf{P}}{2m},$$ (4.6)

$$\mathbf{K}^{(2)} = \left[\frac{\mathbf{P}^4}{8m^3} + \frac{\mathbf{P}^2}{2m^2}\left(\frac{\pi^2}{2\mu} + U^{(0)}(\rho, \pi)\right) + \frac{1}{8}\left(\frac{1}{m_1^3} + \frac{1}{m_2^3}\right)\pi^4\right]\mathbf{Q} - \left(\frac{1}{2}\frac{\mathbf{P}^2}{2m} + \frac{\pi^2}{2\mu} + U^{(0)}(\rho, \pi)\right)\frac{\mathbf{S}\wedge\mathbf{P}}{2m^2} - U^{(1)}(\rho, \pi)\mathbf{Q}, \ldots,$$

with $m = m_1 + m_2$, $\mu = m_1 m_2/m$. Then the Eqs. (3.8), at the various orders in $1/c^2$, take the form

$$-m\frac{\partial\xi_{\tau,j}^{(0)}}{\partial P_i} \equiv \{K_i^{(0)}, x_{\tau,j}^{(0)}\} = 0,$$

$$-m\frac{\partial\xi_{\tau,j}^{(n)}}{\partial P_i} \equiv \{K_i^{(0)}, x_{\tau,j}^{(n)}\} = -\sum_{r=0}^{n-1}\{K_i^{(n-r)}, x_{\tau,j}^{(r)}\} + \sum_{s=0}^{n-1}\sum_{r=0}^{s}x_{\tau,i}^{(n-1-s)}\{x_{\tau,j}^{(s-r)}, H^{(r)}\}, \quad (\tau = 1, 2,; \ n = 1, 2, 3, \cdots).$$ (4.7)

On the other hand the Cauchy conditions (3.13) become in the same way

$$\xi_1^{(n)}(o, \rho, \pi) = \left[\frac{\pi_{20}}{\pi_{10} + \pi_{20}}\right]^{(n)}\rho + \chi_1^{(n)}(\rho, \pi), \quad \xi_2^{(n)}(o, \rho, \pi) = \left[\frac{-\pi_{10}}{\pi_{10} + \pi_{20}}\right]^{(n)}\rho + \chi_2^{(n)}(\rho, \pi),$$ (4.8)

where

$$\frac{\pi_{20}}{\pi_{10} + \pi_{20}} = \sum_{n=0}^{\infty}\frac{1}{c^{2n}}\left[\frac{\pi_{20}}{\pi_{10} + \pi_{20}}\right]^{(n)} = \frac{m_2}{m} + \frac{1}{c^2}\left(\frac{1}{m_2} - \frac{1}{m_1}\right)\frac{\pi^2}{2m} + \frac{1}{c^4}\left[\frac{1}{m_1^2 m_2} + \frac{m_2}{2m}\left(\frac{1}{m_1^3} + \frac{1}{m_2^3}\right) - \frac{1}{m_1 m_2^2} - \frac{1}{2m_2^3}\right]\frac{\pi^4}{4m} + \cdots,$$

$$\frac{-\pi_{10}}{\pi_{10} + \pi_{20}} = \sum_{n=0}^{\infty}\frac{1}{c^{2n}}\left[\frac{-\pi_{10}}{\pi_{10} + \pi_{20}}\right]^{(n)} = -\frac{m_1}{m} + \frac{1}{c^2}\left(\frac{1}{m_2} - \frac{1}{m_1}\right)\frac{\pi^2}{2m} - \frac{1}{c^4}\left[\frac{1}{m_1 m_2^2} + \frac{m_1}{2m}\left(\frac{1}{m_1^3} + \frac{1}{m_2^3}\right) - \frac{1}{m_2 m_1^2} - \frac{1}{2m_1^3}\right]\frac{\pi^4}{4m} + \cdots.$$ (4.9)

Now, for $U(\rho, \pi) \equiv 0$, setting $\chi_1(\rho, \pi) \equiv \chi_2(\rho, \pi) \equiv o$, in agreement with the requirement (c), of Sec. 3, we obtain from Eqs. (4.7)–(4.9),

$$\mathbf{x}_\tau^{(n)} = \mathbf{q}_\tau^{(n)}(\mathbf{Q}, \mathbf{P}, \rho, \pi), \quad \tau = 1, 2,$$ (4.10)

where

$$\mathbf{q}_\tau = \sum_{n=0}^{\infty}(1/c^{2n})\mathbf{q}_\tau^{(n)}, \quad \tau = 1, 2,$$ (4.11)

with [see Eqs. (3.12)]

$$\mathbf{q}_1^{(0)} = \mathbf{Q} + (m_2/m)\rho,$$

$$\mathbf{q}_1^{(1)} = -\frac{m_2}{2m^3}(\rho\cdot\mathbf{P})\mathbf{P} + \left[\left(\frac{1}{m_2} - \frac{1}{m_1}\right)\frac{\pi^2}{2m} - \frac{\pi\cdot\mathbf{P}}{2m^2}\right]\rho + \frac{1}{2m^2}\left(1 - \frac{2m_2}{m_1}\right)(\rho\cdot\mathbf{P})\pi,$$

$$\mathbf{q}_1^{(2)} = \frac{1}{m^3}\left\{\left[\frac{m_2}{m}\left(\frac{3}{4}\frac{\mathbf{P}^2}{2m} + \frac{\pi\cdot\mathbf{P}}{2m_1}\right) + \frac{1}{4}\left(\frac{3}{m_1} - \frac{1}{m_2}\right)\pi^2\right](\rho\cdot\mathbf{P})\mathbf{P} + \left[\left(\frac{1}{4}\frac{\mathbf{P}^2}{2m} + \frac{\pi^2}{2\mu}\right)(\pi\cdot\mathbf{P}) + \frac{m}{8}\left(\frac{m_2}{m_1^3} - \frac{m_1}{m_2^3}\right)\right.\right.$$

$$\left.\left. + \frac{2}{\mu}\left(\frac{1}{m_1} - \frac{1}{m_2}\right)\right)\pi^4\right]\rho + \left[\left(\frac{m_2}{m_1} - \frac{1}{4}\right)\frac{\mathbf{P}^2}{2m} + \frac{1}{2}\left(m_2^2\left(\frac{1}{m_1^3} - \frac{1}{m_2^3}\right) - \frac{3m}{m_1}\left(\frac{1}{m_2} - \frac{1}{m_1}\right)\right)\pi^2 + \frac{m_2}{m_1^2}(\pi\cdot\mathbf{P})\right](\rho\cdot\mathbf{P})\pi\right\}, \ldots,$$ (4.12)

and analogous expressions for $\mathbf{q}_2^{(0)}, \mathbf{q}_2^{(1)}, \mathbf{q}_2^{(2)}$ obtained from Eqs. (3.12) by means of the Z operation (3.16).

In the case of interaction, the same procedure gives

$$\mathbf{x}_1^{(0)} = \mathbf{q}_1^{(0)}, \quad \mathbf{x}_1^{(1)} = \mathbf{q}_1^{(1)} - \frac{m_2^2}{m^3}(\rho\cdot\mathbf{P})\frac{\partial U^{(0)}}{\partial\pi} + \chi_1^{(1)}(\rho, \pi),$$

$$\mathbf{x}_1^{(2)} = \mathbf{q}_1^{(2)} + \frac{1}{m^2}\left\{\frac{m_2}{m^2}(\rho\cdot\mathbf{P})\left[U^{(0)}(\rho, \pi) + \frac{m_2}{2m}\left(\mathbf{P}\cdot\frac{\partial U^{(0)}}{\partial\pi}\right)\right] - \frac{1}{2}(\mathbf{P}\cdot\chi_1^{(1)})\right\}\mathbf{P} + \frac{1}{m^3}\left\{(\pi\cdot\mathbf{P})U^{(0)}(\rho, \pi)\right.$$

$$\left. - \frac{m_2}{4m}(\rho\cdot\mathbf{P})\left(\mathbf{P}\cdot\frac{\partial U^{(0)}}{\partial\rho}\right) + \left(1 - \frac{m_2}{m_1}\right)\left(\pi\cdot\frac{\partial U^{(0)}}{\partial\rho}\right)(\rho\cdot\mathbf{P})\right\}\rho - \frac{1}{m}\left\{\frac{1}{m^2}(\rho\cdot\mathbf{P})\left[\left(1 - \frac{m_2}{m_1}\right)U^{(0)}(\rho, \pi)\right.\right.$$

$$\left.\left. + \frac{m_2}{4m}\left(1 - \frac{4m_2}{m_1}\right)\left(\mathbf{P}\cdot\frac{\partial U^{(0)}}{\partial\pi}\right)\right] + \frac{1}{m_1}(\mathbf{P}\cdot\chi_1^{(1)})\right\}\pi + \left\{\frac{m_2}{4m^4}\left(1 - \frac{2m_2}{m_1}\right)(\rho\cdot\mathbf{P})^2\right\}\frac{\partial U^{(0)}}{\partial\rho} + \frac{m_2}{m^2}\left\{(\rho\cdot\mathbf{P})\left[\frac{m_2}{2m^3}\mathbf{P}^2\right.\right.$$

$$\left.\left. + \frac{1}{4m^2}\left(1 + \frac{4m_2}{m_1}\right)(\pi\cdot\mathbf{P}) + \frac{m_2^2}{m^3}\left(\mathbf{P}\cdot\frac{\partial U^{(0)}}{\partial\pi}\right) + \frac{m_2}{m^2}U^{(0)}(\rho, \pi) - \frac{1}{mm_2}\left(1 - \frac{3m_2}{2m_1}\right)\pi^2\right] - (\mathbf{P}\cdot\chi_1^{(1)})\right\}\frac{\partial U^{(0)}}{\partial\pi}$$ (4.13)

$$+ \frac{m_2^2}{2m^4 m_1} (\rho \cdot P)^2 \left(\sum_k \pi_k \frac{\partial}{\partial \rho_k} \right) \frac{\partial U^{(0)}}{\partial \pi} + \frac{m_2^3}{2m^5} (\rho \cdot P)^2 \left\{ \frac{\partial U^{(0)}}{\partial \pi}, U^{(0)}(\rho, \pi) \right\} - \frac{1}{mm_1} (\rho \cdot P) \left(\pi \cdot \frac{\partial \chi_1^{(1)}}{\partial \rho} \right)$$

$$- \frac{m_2}{m^2} (\rho \cdot P) \{ \chi_1^{(1)}, U^{(0)}(\rho, \pi) \} - \frac{m_2^2}{m^3} (\rho \cdot P) \frac{\partial U^{(1)}}{\partial \pi} + \chi_1^{(2)}(\rho, \pi), \ldots,$$

and analogous expressions for $x_2^{(0)}, x_2^{(1)}, x_2^{(2)}$ obtained from Eqs. (4.13) by means of the $\mathcal{Z}$ operation [recall that $U^{(n)}(\rho, \pi)$ depends on the κ_1, κ_2 in a symmetric way!].

5. "LOCALIZABILITY" AND "CAUSALITY" CONDITIONS FOR THE PARTICLE WORLD LINES: THE ZERO-INTERACTION THEOREM

We shall discuss now to what extent the arbitrariness in the choice of χ_1 and χ_2 in Eq. (4.13) can be exploited for satisfying the "localizability" and "causality" conditions (1.2), (1.3).

We start from the expressions derived in the preceding section for the expansion of the covariant vectors x_1, x_2 and write the Poisson brackets among their various components in the form

$$\{ x_{\tau i}, x_{\tau' j} \} = (1/c^2) \{ x_{\tau i}, x_{\tau' j} \}^{(1)}$$
$$+ (1/c^4) \{ x_{\tau i}, x_{\tau' j} \}^{(2)} + \cdots, \quad \tau, \tau' = 1, 2. \quad (5.1)$$

Let us consider first the "localizability" conditions

$$\{ x_{\tau i}, x_{\tau j} \} = 0, \quad \tau = 1, 2. \quad (5.2)$$

From Eqs. (4.13), at the first order in $1/c^2$ and for $\tau = 1$, we obtain

$$\{ x_{1i}, x_{1j} \}^{(1)}$$
$$= \frac{m_2^2}{m^3} \left(\frac{\partial U^{(0)}}{\partial \pi_i} \rho_j - \frac{\partial U^{(0)}}{\partial \pi_j} \rho_i \right) - \frac{m_2}{m} \left(\frac{\partial \chi_{1i}^{(1)}}{\partial \pi_j} - \frac{\partial \chi_{1j}^{(1)}}{\partial \pi_i} \right)$$
$$= - \frac{m_2^2}{m^3} \frac{1}{\pi} \frac{\partial U^{(0)}}{\partial \pi} (\rho_i \pi_j - \rho_j \pi_i) - \frac{m_2}{m} \left(\frac{\partial \chi_{1i}^{(1)}}{\partial \pi_j} - \frac{\partial \chi_{1j}^{(1)}}{\partial \pi_i} \right). \quad (5.3)$$

The rhs of these equations is identically zero if we choose

$$\chi_{1i}^{(1)} = - \frac{m_2}{m^2} U^{(0)}(\rho, \pi) \rho_i + \frac{\partial \lambda_1^{(1)}(\rho, \pi)}{\partial \pi_i}, \quad (5.4)$$

where $\lambda_1^{(1)}$ is an arbitrary function of ρ, π, σ subjected only to the general requirements (a)–(d) of Sec. 3; in particular $\lambda_1^{(1)}$ must be odd in σ. In a similar way the "localizability" conditions at first order for the particle 2 give

$$\chi_{2i}^{(1)} = \frac{m_1}{m^2} U^{(0)}(\rho, \pi) \rho_i + \frac{\partial \lambda_2^{(1)}(\rho, \pi)}{\partial \pi_i}. \quad (5.5)$$

It is then easy to see that the "localizability" conditions for each particle can be satisfied to any order in $1/c^2$. In fact, in the center-of-mass system, we must have, for any n, by induction

$$\{ x_{1i}, x_{1j} \}^{(n)} |_{P=0} = f_1^{(n)}(\rho, \pi, \sigma)(\rho_i \pi_j - \rho_j \pi_i)$$
$$- \frac{m_2}{m} \left(\frac{\partial \chi_{1i}^{(n)}}{\partial \pi_j} - \frac{\partial \chi_{1j}^{(n)}}{\partial \pi_i} \right), \quad (5.6)$$

where $f_1^{(n)}(\rho, \pi, \sigma)$ contains the choices previously made for $\chi_{1i}^{(n-1)}, \ldots, \chi_{1i}^{(1)}$. Equation (5.6) follows from the fact that $\rho_i \pi_j - \rho_j \pi_i$ is the only antisymmetrical tensor of rank two which can be constructed out of ρ and π. The equation

$$\frac{\partial \chi_{1i}^{(n)}}{\partial \pi_j} - \frac{\partial \chi_{1j}^{(n)}}{\partial \pi_i} = \frac{m}{m_2} f_1^{(n)}(\rho, \pi, \sigma)(\rho_i \pi_j - \rho_j \pi_i) \quad (5.7)$$

is obviously integrable and can actually be integrated explicitly provided one succeeds in rearranging $(m/m_2) f_1^{(n)}$ in the form

$$\frac{m}{m_2} f_1^{(n)}(\rho, \pi, \sigma) = \frac{1}{\pi} \frac{\partial}{\partial \pi} \mathcal{J}_1^{(n)}(\rho, \pi, \sigma) - \frac{\partial}{\partial \sigma} \mathcal{G}_1^{(n)}(\rho, \pi, \sigma), \quad (5.8)$$

since in this case the general integral is simply

$$\chi_1^{(n)} \equiv \chi_1^{(n)*} + \frac{\partial \lambda_1^{(n)}}{\partial \pi} = \mathcal{J}_1^{(n)} \rho + \mathcal{G}_1^{(n)} \pi + \frac{\partial \lambda_1^{(n)}}{\partial \pi}. \quad (5.9)$$

Of course any free shift of terms between $\mathcal{J}_1^{(n)}$ and $\mathcal{G}_1^{(n)*}$ in Eq. (5.8) modifies the particular determination $\chi_1^{(n)}$ by a π gradient. Repeating the same calculations for $\tau = 2$ we see that $\chi_\tau^{(1)}, \chi_\tau^{(2)}, \cdots$ can be chosen in such a way that the "localizability" conditions are satisfied to any order in the center-of-mass system. On the other hand let us consider the equations

$$\{ K_i, \{ x_{\tau j}, x_{\tau h} \} \} = (1/c^2) [\{ x_{\tau j}, x_{\tau i} \} \{ x_{\tau h}, H \} - \{ x_{\tau h}, x_{\tau i} \} \{ x_{\tau j}, H \}$$
$$+ x_{\tau i} \{ H, \{ x_{\tau j}, x_{\tau h} \} \}], \quad (5.10)$$

which follow from Eq. (3.4) and the Jacobi identity. These relations imply that the "localizability" conditions are satisfied in any reference frame provided that they are in a particular one. Therefore, if Eqs. (5.7) hold true for every n, the "localizability" conditions are verified also in the laboratory frame. For $n = 2$ we can write for example

$$\{ x_{1i}, x_{1j} \}^{(2)} |_{P=0}$$
$$= \frac{1}{\pi} \frac{\partial}{\partial \pi} \left[\frac{1}{m^3} \left(\frac{2m_2}{m_1} - 1 \right) \frac{\pi^2}{2} U^{(0)}(\rho, \pi) + \frac{m_2^2}{m^4} (U^{(0)}(\rho, \pi))^2 \right.$$
$$- \frac{m_2^2}{m^3} U^{(1)}(\rho, \pi) - \frac{1}{m_2 m} \left(\frac{\sigma}{\rho} \frac{\partial \lambda_1^{(1)}}{\partial \rho} + \pi^2 \frac{\partial \lambda_1^{(1)}}{\partial \sigma} \right) + \frac{1}{mm_2}$$
$$\times \frac{\pi^2}{2} \frac{\partial \lambda_1^{(1)}}{\partial \sigma} - \frac{m_2}{m^2} \frac{\partial \lambda_1^{(1)}}{\partial \sigma} U^{(0)}(\rho, \pi) \right] + \frac{\partial}{\partial \sigma} \left[\frac{m_2}{m^2} \frac{1}{\pi} \frac{\partial \lambda_1^{(1)}}{\partial \pi} \right.$$
$$\times U^{(0)}(\rho, \pi) - \frac{1}{mm_2} \frac{\pi^2}{2} \frac{1}{\pi} \frac{\partial \lambda_1^{(1)}}{\partial \pi} - \frac{1}{m_1 m} \lambda_1^{(1)} \right]$$
$$\times (\rho_i \pi_j - \rho_j \pi_i) - \frac{m_2}{m} \left(\frac{\partial \chi_{1i}^{(2)}}{\partial \pi_j} - \frac{\partial \chi_{1j}^{(2)}}{\partial \pi_i} \right) \quad (5.11)$$

and the general solution for $\chi_1^{(2)}(\rho, \pi)$ results correspondingly, i.e.,

$$\chi_1^{(2)} = \left[\frac{1}{m_2 m^2} \left(\frac{2m_2}{m} - 1 \right) \frac{\pi^2}{2} U^{(0)}(\rho, \pi) + \frac{m_2}{m^3} (U^{(0)}(\rho, \pi))^2 \right.$$
$$\left. - \frac{m_2}{m^2} U^{(1)}(\rho, \pi) \right] \rho + \left[- \frac{1}{m} \frac{\partial \lambda_1^{(1)}}{\partial \pi} U^{(0)}(\rho, \pi) \right.$$
$$- \frac{1}{m_2^2} \left(\pi \cdot \frac{\partial \lambda_1^{(1)}}{\partial \rho} \right) \rho + \frac{1}{2m_2^2} \pi^2 \frac{\partial \lambda_1^{(1)}}{\partial \pi} + \frac{1}{m_1 m_2} \lambda_1^{(1)}(\rho, \pi) \pi$$
$$- \frac{m}{2m_2} \left\{ \lambda_1^{(1)}, \frac{\partial \lambda_1^{(1)}}{\partial \pi} \right\} + \frac{\partial \lambda_1^{(2)}}{\partial \pi}, \quad (5.12)$$

while the expression for $\chi_2^{(2)}(\rho, \pi)$ can be obtained from

(5.11) and (5.12) by means of the Z operation [which in particular implies $\lambda_1^{(1)} \doteq -\lambda_2^{(1)}$ due to Eqs. (3.14), (3.17), (5.4), (5.5)].

It is clear from Eqs. (5.4), (5.5), (5.7), (5.11), (5.12) that this procedure determines $\chi_1^{(n)}(\rho, \pi)$ and $\chi_2^{(n)}(\rho, \pi)$ up to arbitrary [apart from the conditions (a)—(d)] additive gradients

$$
\chi_1 = \frac{1}{c^2}\left[\chi_1^{(1)*}(\rho,\pi) + \frac{\partial \lambda_1^{(1)}}{\partial \pi}\right]
$$
$$
+ \frac{1}{c^4}\left[\chi_1^{(2)*}(\rho,\pi \mid \lambda_1^{(1)}) + \frac{\partial \lambda_1^{(2)}}{\partial \pi}\right] + \cdots,
$$

$$
\chi_2 = \frac{1}{c^2}\left[\chi_2^{(1)*}(\rho,\pi) + \frac{\partial \lambda_2^{(1)}}{\partial \pi}\right]
$$
$$
+ \frac{1}{c^4}\left[\chi_2^{(2)*}(\rho,\pi \mid \lambda_2^{(1)}) + \frac{\partial \lambda_2^{(2)}}{\partial \pi}\right] + \cdots,
$$

(5.13)

being $\chi_1^{(n)*}, \chi_2^{(n)*}$, particular determinations of $\chi_1^{(n)}, \chi_2^{(n)}$. We shall write shortly

$$
\chi_1 = \chi_1^* + \frac{\partial \lambda_1}{\partial \pi}, \quad \chi_2 = \chi_2^* + \frac{\partial \lambda_2}{\partial \pi}, \tag{5.14}
$$

keeping in mind, however, that $\chi_r^{(n)*}$ depends on the choices of $\lambda_r^{(1)}, \lambda_r^{(2)}, \dots, \lambda_r^{(n-1)}$! Finally it is convenient to introduce the functions

$$
\lambda = \lambda_1 - \lambda_2, \quad \Lambda = (m_1\lambda_1 + m_2\lambda_2)/m, \tag{5.15}
$$

or

$$
\lambda_1 = \Lambda + (m_2/m)\lambda, \quad \lambda_2 = \Lambda - (m_1/m)\lambda. \tag{5.16}
$$

Then [see Eqs. (3.23), (3.24)] we have

$$
\chi = \chi^* + \frac{\partial \lambda}{\partial \pi},
$$

$$
\mathcal{X} = \mathcal{X}^* + \frac{\partial \Lambda}{\partial \pi} + \frac{m_2\pi_{10} - m_1\pi_{20}}{m(\pi_{10} + \pi_{20})}\frac{\partial \lambda}{\partial \pi},
$$

(5.17)

and we can write

$$
\mathbf{x}_{1_{c.m.}} - \mathbf{x}_{2_{c.m.}} = \rho + \chi^* + \frac{\partial \lambda}{\partial \pi}, \tag{5.18}
$$

$$
\frac{\pi_{10}\mathbf{x}_{1_{c.m.}} + \pi_{20}\mathbf{x}_{2_{c.m.}}}{\pi_{10} + \pi_{20}}
$$
$$
= Q + \mathcal{X}^* + \frac{\partial \Lambda}{\partial \pi} + \frac{m_2\pi_{10} - m_1\pi_{20}}{m(\pi_{10} + \pi_{20})}\frac{\partial \lambda}{\partial \pi}. \tag{5.19}
$$

We want to show now that the arbitrary functions λ and Λ play very different roles. As we have already pointed out, the explicit choice of the vectors $\chi_1(\rho,\pi)$, $\chi_2(\rho,\pi)$ [or the functions $\lambda(\rho,\pi)$, $\Lambda(\rho,\pi)$] must be considered as a part of a particular dynamical theory together with the choice of the canonical interaction potential $U(\rho,\pi)$. On the other hand, if we take an opposite point of view and assume we have assigned expressions of the canonical generators of the Poincaré group in terms of *physical coordinates* x_1, x_2 and v_1, v_2 and assign Poisson bracket relations among these variables, the internal vectors $\rho(x_1,x_2,v_1,v_2)$ and $\pi(x_1,x_2,v_1,v_2)$ are obviously determined up to a global canonical equivalence which preserves their "internal" character and Euclidean transformation properties (see Sec. 2) and is compatible with the asymptotic condition (c). (We sketch in the Appendix the structure of this inverse problem which in turn sheds further light into the physical meaning of the theory in the canonical form.) Clearly, any canonical transformation of this kind amounts to a proper redefinition of the vectors $\chi_1(\rho,\pi)$ and $\chi_2(\rho,\pi)$ [or $\lambda(\rho,\pi)$, $\Lambda(\rho,\pi)$] and of the canonical potential $U(\rho,\pi)$. Therefore *there must exist classes of different assignments of the functions* $U(\rho,\pi)$, $\lambda(\rho,\pi)$, $\Lambda(\rho,\pi)$ *which are physically equivalent.* As a matter of fact we will show presently that *only two among these functions are physically essential* for the dynamical description of the interacting particles. More precisely, we are going to see that *the arbitrariness connected with the choice of the function* $\lambda(\rho,\pi)$ *in Eqs.* (5.18), (5.19) *can be removed by means of a canonical redefinition of the internal vectors* ρ, π *and of the canonical potential* $U(\rho,\pi)$.

The most simple way to exploit this fact is a step-wise renormalization to be performed order-by-order. Up to the first order in $1/c^2$ we can write

$$
\mathbf{x}_{1_{c.m.}} - \mathbf{x}_{2_{c.m.}} = \rho + \frac{1}{c^2}\left[\chi^{(1)*} + \frac{\partial \lambda^{(1)}}{\partial \pi}\right] + \cdots \tag{5.20}
$$

where the "localizability" conditions have been satisfied at this order with certain determinations $\chi_1^{(1)*}, \chi_2^{(1)*}$. Then, performing the canonical transformation

$$
T^{(1)}: \begin{cases} \rho' = \exp\left[-\frac{1}{c^2}\{\lambda^{(1)}, \cdots\}\right]\rho = \rho + \frac{1}{c^2}\frac{\partial \lambda^{(1)}}{\partial \pi} + \frac{1}{c^4}\frac{1}{2}\left\{\frac{\partial \lambda^{(1)}}{\partial \pi}, \lambda^{(1)}\right\} + \cdots, \\ \pi' = \exp\left[-\frac{1}{c^2}\{\lambda^{(1)}, \cdots\}\right]\pi = \pi - \frac{1}{c^2}\frac{\partial \lambda^{(1)}}{\partial \rho} - \frac{1}{c^4}\frac{1}{2}\left\{\frac{\partial \lambda^{(1)}}{\partial \rho}, \lambda^{(1)}\right\} + \cdots, \end{cases} \tag{5.21}
$$

and introducing correspondingly the renormalized canonical potential

$$
U'(\rho',\pi') = Mc^2 - c\sqrt{m_1^2 c^2 + \pi'^2} - c\sqrt{m_2^2 c^2 + \pi'^2}
$$
$$
= U^{(0)}(\rho',\pi') + \frac{1}{c^2}\left[U^{(1)}(\rho',\pi') + \{\lambda^{(1)}(\rho',\pi'), U^{(0)}(\rho',\pi')\} + \frac{1}{\mu}\pi' \cdot \frac{\partial \lambda^{(1)}}{\partial \rho'}(\rho',\pi')\right] + \cdots
$$
$$
= U^{(0)}(\rho',\pi') + \frac{1}{c^2}[U^{(1)}(\rho',\pi') + \{\lambda^{(1)}(\rho',\pi'), (Mc^2)^{(0)}\}] + \cdots, \tag{5.22}
$$

with

$$
(Mc^2)^{(0)} = H^{(0)}(\rho',\pi',o) = U^{(0)}(\rho',\pi') + \frac{\pi'^2}{2\mu}, \tag{5.23}
$$

Eq. (5.20) becomes

$$\mathbf{x}_{1_{c.m.}} - \mathbf{x}_{2_{c.m.}} = \rho' + \frac{1}{c^2}\chi^{(1)*}(\rho', \pi') + O\left(\frac{1}{c^4}\right). \tag{5.24}$$

In this way, due to the structure of Eq. (5.19), we have eliminated $\lambda^{(1)}(\rho, \pi)$ from the expressions of $\mathbf{x}_{1_{c.m.}}$ and $\mathbf{x}_{2_{c.m.}}$ separately, up to the first order in $1/c^2$ and therefore, via Lorentz transformation, from the expressions of $\mathbf{x}_1(\mathbf{Q}, \mathbf{P}, \rho, \pi)$ and $\mathbf{x}_2(\mathbf{Q}, \mathbf{P}, \rho, \pi)$ at the same order. At this point we can use these new expressions to solve the "localizability" conditions at the order $1/c^4$. Denoting by $\chi^{(2)*'}$ the corresponding new choice of $\chi^{(2)*}$, we can write

$$\mathbf{x}_{1_{c.m.}} - \mathbf{x}_{2_{c.m.}} = \rho' + \frac{1}{c^2}\chi^{(1)*}(\rho', \pi') + \frac{1}{c^4}\left[\chi^{(2)*'}(\rho', \pi') + \frac{\partial\lambda^{(2)}}{\partial\pi'}\right] + \cdots. \tag{5.25}$$

Then by performing the canonical transformation

$$T^{(2)}: \begin{cases} \rho'' = \exp\left[-\frac{1}{c^4}\{\lambda^{(2)}, \cdots\}\right]\rho' = \rho' + \frac{1}{c^4}\frac{\partial\lambda^{(2)}}{\partial\pi'} + O\left(\frac{1}{c^8}\right), \\[2mm] \pi'' = \exp\left[-\frac{1}{c^4}\{\lambda^{(2)}, \cdots\}\right]\pi' = \pi' - \frac{1}{c^4}\frac{\partial\lambda^{(2)}}{\partial\rho'} + O\left(\frac{1}{c^8}\right), \end{cases} \tag{5.26}$$

and introducing the renormalized second order canonical potential

$$U''(\rho'', \pi'') = U^{(0)}(\rho'', \pi'') + \frac{1}{c^2}U^{(1)'}(\rho'', \pi'') + \frac{1}{c^4}[U^{(2)}(\rho'', \pi'') + \{\lambda^{(2)}(\rho'', \pi''), (Mc^2)^{(0)}\}$$

$$+ \{\lambda^{(1)}(\rho'', \pi''), (Mc^2)^{(1)}\} + \tfrac{1}{2}\{\lambda^{(1)}(\rho'', \pi''), \{\lambda^{(1)}(\rho'', \pi''), (Mc^2)^{(0)}\}\}] + \cdots, \tag{5.27}$$

with

$$(Mc^2)^{(1)} = H^{(1)}(\rho'', \pi'', 0) = -\frac{1}{8}\left(\frac{1}{m_1^3} + \frac{1}{m_2^3}\right)\pi''^4 + U^{(1)}(\rho'', \pi''), \tag{5.28}$$

Eq. (5.25) takes the form

$$\mathbf{x}_{1_{c.m.}} - \mathbf{x}_{2_{c.m.}} = \rho'' + \frac{1}{c^2}\chi^{(1)*}(\rho'', \pi'') + \frac{1}{c^4}\chi^{(2)*'}(\rho'', \pi'') + O\left(\frac{1}{c^6}\right), \tag{5.29}$$

and, due to the structure of Eq. (5.19), we have succeeded in eliminating the function $\lambda^{(2)}(\rho, \pi)$ from the expressions of $\mathbf{x}_{1_{c.m.}}$, $\mathbf{x}_{2_{c.m.}}$, i.e., of $\mathbf{x}_1$, $\mathbf{x}_2$ at the second order. By iterating this step-wise procedure it is seen that by means of the product canonical transformation $T = \cdots T^{(n)} \cdots T^{(2)} \cdot T^{(1)}$, i.e., by means of a redefinition of the internal vectors ρ, π together with the introduction of a renormalized canonical potential $\tilde{U}(\rho, \pi)$, the function $\lambda(\rho, \pi)$ can always be disposed of.

Let us consider now the "causality" conditions [Eqs. (5.1) for $\tau' \neq \tau$] assuming that the "localizability" conditions are satisfied up to any order. From Eqs. (4.13), at the first order in $1/c^2$, we obtain via Eqs. (5.4), (5.5), (5.15), (5.16),

$$\{x_{1i}, x_{2j}\}^{(1)} = \frac{m_2^2}{m^3}\rho_j\frac{\partial U^{(0)}}{\partial\pi_i} - \frac{m_1^2}{m^3}\rho_j\frac{\partial U^{(0)}}{\partial\pi_j} - \frac{\mu}{m^2}(\rho \cdot \mathbf{P})\frac{\partial^2 U^{(0)}}{\partial\pi_i\partial\pi_j} + \frac{m_2}{m}\frac{\partial\chi_{2j}^{(1)}}{\partial\pi_i} + \frac{m_1}{m}\frac{\partial\chi_{1i}^{(1)}}{\partial\pi_j}$$

$$= \frac{m_2}{m^2}\rho_i\frac{\partial U^{(0)}}{\partial\pi_i} - \frac{m_1}{m^2}\rho_i\frac{\partial U^{(0)}}{\partial\pi_j} - \frac{\mu}{m^2}(\rho \cdot \mathbf{P})\frac{\partial^2 U^{(0)}}{\partial\pi_i\partial\pi_j} + \frac{\partial^2\Lambda^{(1)}}{\partial\pi_i\partial\pi_j}. \tag{5.30}$$

We ask now if there is any situation in which the rhs of Eqs. (5.30) vanishes identically. Let us observe first that the "causality" conditions are not automatically satisfied in all of the reference frames if they are in a particular one, unlike the "localizability" conditions. Actually, this can be easily seen from the relation

$$\{K_i, \{x_{1j}, x_{2h}\}\} = x_{2i}\{\{x_{1j}, x_{2h}\}, H\} + (x_{1i} - x_{2j})\{x_{2h}, \dot{x}_{1j}\}, \tag{5.31}$$

since the Poisson bracket $\{x_{2h}, \dot{x}_{1j}\}$ fails to vanish already at zeroth order. Then, the rhs of Eqs. (5.30) vanishes identically if and only if

$$\frac{m_2}{m^2}\rho_j\frac{\partial U^{(0)}}{\partial\pi_i} - \frac{m_1}{m^2}\rho_i\frac{\partial U^{(0)}}{\partial\pi_j} + \frac{\partial^2\Lambda^{(1)}}{\partial\pi_i\partial\pi_j} = 0 \tag{5.32}$$

and

$$\frac{\partial^2 U^{(0)}}{\partial\pi_i\partial\pi_j} = 0. \tag{5.33}$$

This last equation implies that $U^{(0)}$ has the form

$$U^{(0)}(\rho, \pi, \sigma) = U^{(0)}(\rho) + U_1^{(0)}(\rho) \cdot \sigma, \tag{5.34}$$

while the invariance of $U(\rho, \pi)$ under the anticanonical time reflection (see Sec. 2), imposes

$$U_1^{(0)}(\rho) \equiv 0,$$

i.e.,

$$U^{(0)}(\rho,\pi) \equiv U^{(0)}_{[\kappa_1,\kappa_2]}(\rho).$$ (5.36)

Finally, inserting Eq. (5.36) into Eq. (5.22) and taking into account the fact that $\Lambda^{(1)}$ must be an odd function of σ in force of Eq. (3.6) [see Eqs. (3.14), (3.15), (3.24), (5.15), (5.17)], we can write

$$\Lambda^{(1)}(\rho,\pi,\sigma) \equiv (1/2m)\Omega^{(1)}_{[\kappa_1,\kappa_2]}(\rho) \cdot \sigma,$$ (5.37)

where $\Omega^{(1)}_{[\kappa_1,\kappa_2]}$ must be an antisymmetrical function of the parameters κ_1, κ_2 in order to satisfy the *symmetry condition* (b) of Sec. 3.

Let us consider now the "causality" conditions at the order $1/c^4$. We shall assume that the canonical transformation $T^{(1)}$ [Eq. (5.21)] has been carried out so that we can put

$$\lambda^{(1)}(\rho,\pi) \equiv 0$$ (5.38)

and the canonical potential U is actually the *renormalized potential* up to this order [Eqs. (5.27), (5.22)]. The "causality" conditions can be calculated using Eqs. (4.13) together with Eqs. (5.4), (5.5), (5.12), (5.16), (5.36), (5.37), (5.38). We obtain

$$\begin{aligned}
\{x_{1i}, x_{2j}\}^{(2)} &= \frac{1}{m^2}\left\{(\rho\cdot\mathbf{P})\left[\frac{1}{m}U^{(0)}(\rho) + \frac{1}{2}\left(\frac{1}{m_1}-\frac{1}{m_2}\right)\Omega^{(1)}(\rho)\right]\right\}\delta_{ij} + \frac{1}{m^2}\left\{(\rho\cdot\mathbf{P})\left[\frac{1}{m^2}\left(\frac{m_1^2}{m_2}+\frac{m_2^2}{m_1}+2m\right)\frac{1}{\rho}\frac{dU^{(0)}}{d\rho}\right.\right. \\
&\quad \left.+\frac{1}{2}\left(\frac{1}{m_1}-\frac{1}{m_2}\right)\frac{1}{\rho}\frac{d\Omega^{(1)}}{d\rho}\right] + (\rho\cdot\pi)\left[\left(\frac{1}{m_1}-\frac{1}{m_2}\right)\frac{1}{\rho}\frac{dU^{(0)}}{d\rho} + \left(\frac{1}{2\mu}-\frac{m_1}{m_2^2}-\frac{m_2}{m_1^2}\right)\frac{1}{\rho}\frac{d\Omega^{(1)}}{d\rho}\right]\right\}\rho_i\rho_j \\
&\quad + \frac{1}{m^2}\left\{\frac{1}{m_2}U^{(0)}(\rho)+\frac{m_1}{2\mu}\left(\frac{1}{m_1}-\frac{1}{m_2}\right)\Omega^{(1)}(\rho)\right\}\rho_i\pi_j + \frac{1}{m^2}\left\{-\frac{1}{m_1}U^{(0)}(\rho)-\frac{m_2}{2\mu}\left(\frac{1}{m_1}-\frac{1}{m_2}\right)\Omega^{(1)}(\rho)\right\}\rho_j\pi_i \\
&\quad - \frac{m_1^2}{m^3}\rho_i\frac{\partial U^{(1)}}{\partial\pi_j} + \frac{m_2^2}{m^3}\rho_j\frac{\partial U^{(1)}}{\partial\pi_i} - \frac{\mu}{m^2}(\rho\cdot\mathbf{P})\frac{\partial^2 U^{(1)}}{\partial\pi_i\partial\pi_j} + \frac{\partial\Lambda^{(2)}}{\partial\pi_i\partial\pi_j}.
\end{aligned}$$ (5.39)

In order that the rhs of Eqs. (5.39) vanishes identically, it is necessary and sufficient that the coefficients of the various independent tensors vanish separately, in particular for $\mathbf{P} = 0$. The vanishing condition for the coefficient of δ_{ij}, $\frac{1}{2}(\rho_i\pi_j+\rho_j\pi_i)$, $\frac{1}{2}(\rho_i\pi_j-\rho_j\pi_i)$, provides a system of linear homogeneous equations in the expressions $U^{(0)}(\rho)$, $\Omega^{(1)}(\rho)$, $(1/\pi)(\partial U^{(1)}/\partial\pi)$, $(1/\pi)\times(\partial^2\Lambda^{(2)}/\partial\pi\,\partial\sigma)$. Eliminating the last two expressions, we obtain the conditions

$$U^{(0)}(\rho) \equiv 0, \quad \Omega^{(1)}(\rho) \equiv 0.$$ (5.40)

Then, the causality conditions at the second order take the form

$$\frac{m_2}{m^2}\rho_j\frac{\partial U^{(1)}}{\partial\pi_i} - \frac{m_1}{m^2}\rho_i\frac{\partial U^{(1)}}{\partial\pi_j} + \frac{\partial^2\Lambda^{(2)}}{\partial\pi_i\,\partial\pi_j} = 0,$$ (5.41)

$$\frac{\partial^2 U^{(1)}}{\partial\pi_i\,\partial\pi_j} = 0,$$ (5.42)

which are identical to Eqs. (5.32), (5.33) apart from the replacement of $U^{(0)}, \Lambda^{(1)}$ with $U^{(1)}, \Lambda^{(2)}$, respectively, and give

$$U^{(1)}(\rho,\pi) = U^{(1)}_{[\kappa_1,\kappa_2]}(\rho),$$ (5.43)

$$\Lambda^{(2)}(\rho,\pi) = (1/2m)\Omega^{(2)}_{[\kappa_1,\kappa_2]}(\rho)\cdot\sigma$$ (5.44)

$$[\Omega^{(2)}_{[\kappa_1,\kappa_2]}(\rho) = -\Omega^{(2)}_{[\kappa_2,\kappa_1]}(\rho)].$$

The result can be easily extended to higher orders in $1/c^2$ in the following way: If we assume that the "localizability" conditions are satisfied up to any order and the corresponding canonical transformations $T^{(1)}, T^{(2)}, \ldots, T^{(n-1)}$ have been carried out, the requirement that the "causality" conditions are satisfied up to the order n implies

$$U^{(0)} \equiv \tilde{U}^{(1)} \equiv \cdots \equiv \tilde{U}^{(n-2)} \equiv 0,$$ (5.45)

$$\Lambda^{(1)} \equiv \Lambda^{(2)} \equiv \cdots \equiv \Lambda^{(n-1)} \equiv 0,$$ (5.46)

$$\tilde{U}^{(n-1)}(\rho,\pi) = \tilde{U}^{(n-1)}_{[\kappa_1,\kappa_2]}(\rho),$$ (5.47)

$$\Lambda^{(n)}(\rho,\pi) = (1/2m)\Omega^{(n)}_{[\kappa_1,\kappa_2]}(\rho)\cdot\sigma,$$ (5.48)

where $\tilde{U}^{(k)}$ means the kth order (in $1/c^2$) potential renormalized by the product transformation $T^{(n-1)}\cdots T^{(2)}\cdot T^{(1)}$. The validity of such generalization can be easily proved by induction. Actually if we write

$$\mathbf{x}_\tau^{(n)} = \mathbf{q}_\tau^{(n)} + \mathbf{\Psi}_\tau^{(n)}, \quad (\tau=1,2),$$ (5.49)

and assume the above statement to hold true up to the order $n-1$, we can easily see that we must have

$$\mathbf{\Psi}_\tau^{(0)} \equiv \mathbf{\Psi}_\tau^{(1)} \equiv \cdots \equiv \mathbf{\Psi}_\tau^{(n-2)} \equiv 0, \quad (\tau=1,2),$$ (5.50)

while the expressions

$$\mathbf{\Psi}_\tau^{(n-1)}, \mathbf{\Psi}_\tau^{(n)}, \quad (\tau=1,2),$$ (5.51)

are obtained from the corresponding expressions for $n=2$ [see Eqs. (4.13) together with (5.12), (5.16), (5.36)–(5.38), (5.43), (5.44)] by means of the following substitutions:

$$U^{(0)} \to \tilde{U}^{(n-2)}, \quad U^{(1)} \to \tilde{U}^{(n-1)},$$
$$\Omega^{(1)} \to \Omega^{(n-1)}, \quad \Lambda^{(2)} \to \Lambda^{(n)}.$$ (5.52)

Therefore the rhs of the equations

$$\begin{aligned}
\{x_{1i}, x_{2j}\}^{(n)} &= \{q_{1i}^{(0)}, \mathbf{\Psi}_{2j}^{(n)}\} + \{\mathbf{\Psi}_{1i}^{(n)}, q_{2j}^{(0)}\} \\
&\quad + \{q_{1i}^{(1)}, \mathbf{\Psi}_{2j}^{(n-1)}\} + \{\mathbf{\Psi}_{1i}^{(n-1)}, q_{2j}^{(1)}\}
\end{aligned}$$ (5.53)

is derived from the rhs of Eqs. (5.39) by means of the same replacements and the statement is established up to the order n. [Note that at the order $n=2$ one has

$$\begin{aligned}
\{x_{1i}, x_{2j}\}^{(2)} &= \{q_{1i}^{(0)}, \mathbf{\Psi}_{2j}^{(2)}\} + \{\mathbf{\Psi}_{1i}^{(2)}, q_{2j}^{(0)}\} \\
&\quad + \{q_{1i}^{(1)}, \mathbf{\Psi}_{2j}^{(1)}\} + \{\mathbf{\Psi}_{1i}^{(1)}, q_{2j}^{(1)}\} + \{\mathbf{\Psi}_{1i}^{(1)}, \mathbf{\Psi}_{2j}^{(1)}\},
\end{aligned}$$

but if Eqs. (5.36), (5.37) are satisfied, the last term is identically zero.]

In conclusion it is clear that the "causality" conditions can be satisfied up to any order if and only if the canoni-

cal potential is identically zero. If analyticity in $1/c^2$ is assumed for the theory, this appears in turn as a new proof of the zero-interaction theorem. [Note that the impossibility of satisfying the "causality" conditions does not depend on the enforcement of the "localizability" conditions. It is interesting also to make a comparison with the conditions imposed by Bel,[13] i.e., $\{x_{1i} - x_{2i}, x_{1j} - x_{2j}\} = 0$. Assuming the validity of Eqs. (1.2), these conditions amount to requiring that the antisymmetrical part (in i, j) of Eqs. (1.3) be identically satisfied. It is seen, e.g., at the order $1/c^4$, that these conditions are not satisfied in general (see also the physically significant cases corresponding to the restrictions (6.54).]

We have seen so far that, after a suitable redefinition of the internal variables, an arbitrary function $\Lambda(\rho, \pi)$ is left in the expressions of the physical position variables x_1, x_2 once the "localizability" conditions have been enforced. Even if we cannot dispose of this arbitrariness in order to impose the "causality" conditions, we can however try to reduce the deviations of the theory from a strict "causality" as much as possible. This is actually feasible in the following particular sense. We can write in the center-of-mass system at any order in $1/c^2$

$$\{x_{1i}, x_{2j}\}^{(n)} \big|_{P=0} = \gamma_{ij}^{(n)}(\rho, \pi) + \frac{\partial^2 \Lambda^{(n)}}{\partial \pi_i \, \partial \pi_j} , \qquad (5.54)$$

where $\gamma_{ij}^{(n)}(\rho, \pi)$ depends on $U^{(0)}, \tilde{U}^{(1)}, \ldots, \tilde{U}^{(n-1)}$ and $\Lambda^{(1)}, \Lambda^{(2)}, \ldots, \Lambda^{(n-1)}$. Then we can write

$$\gamma_{ij}^{(n)} = [\gamma_{ij}^{(n)}]^{II} + [\gamma_{ij}^{(n)}]^{IS} + [\gamma_{ij}^{(n)}]^{SI} + [\gamma_{ij}^{(n)}]^{SS},$$

where $[\gamma_{ij}^{(n)}]^{II}$ denotes a tensor which is irrotational in both indices for what concerns the variables π; $[\gamma_{ij}^{(n)}]^{IS}$ is a tensor which is irrotational in the first index and solenoidal in the second index for the same variables, and so on. The term $[\gamma_{ij}^{(n)}]^{II}$ can always be assumed to be symmetrical in i, j since it can be shown that its possible antisymmetrical part must be independent of π and thus it can be absorbed in other terms. Then $\Lambda^{(n)}$ can always be chosen in such a way to cancel out the term $[\gamma_{ij}^{(n)}]^{II}$. Once this condition has been imposed, $\Lambda^{(n)}$ is just determined up to a term having the form $(1/2m)\Omega^{(n)}(\rho) \cdot \sigma$ which amounts to a shift parallel to ρ in the expression of the "kinematical" center of mass (5.19).

Before concluding this section it is worthwhile to note that it is also possible to give a compact procedure for the canonical redefinition of the internal variables introduced before on the basis of a step-wise method. First of all it can be shown that, after having solved the "localizability" conditions once for all at any order, the expressions $\chi^{(n)*}(\rho, \pi), \mathcal{X}^{(n)*}(\rho, \pi)$ can be given a typical structure of the form

$$\chi^{(n)*}(\rho, \pi) = \chi_A^{(n)*}[\rho, \pi \,|\, \tilde{U}^{(0)}, \ldots, \tilde{U}^{(n-1)} \,|\, \Lambda^{(1)}, \ldots, \Lambda^{(n-1)}]$$
$$+ \chi_B^{(n)*}[\rho, \pi \,|\, \tilde{U}^{(0)}, \ldots, \tilde{U}^{(n-1)} \,|\, \Lambda^{(1)},$$
$$\ldots, \Lambda^{(n-1)} \,|\, \lambda^{(1)}, \ldots, \lambda^{(n-1)}] \qquad (5.56)$$

$$\mathcal{X}^{(n)*}(\rho, \pi) = \mathcal{X}_A^{(n)*}[\rho, \pi \,|\, \tilde{U}^{(0)}, \ldots, \tilde{U}^{(n-1)} \,|\, \Lambda^{(1)}, \ldots, \Lambda^{(n-1)}]$$
$$+ \mathcal{X}_B^{(n)*}[\rho, \pi \,|\, \tilde{U}^{(0)}, \ldots, \tilde{U}^{(n-1)} \,|\, \Lambda^{(1)},$$
$$\ldots, \Lambda^{(n-1)} \,|\, \lambda^{(1)}, \ldots, \lambda^{(n-1)}], \qquad (5.57)$$

where

$$\tilde{U}^{(0)} = U^{(0)}, \quad \tilde{U}^{(1)} = U^{(1)} + \{\lambda^{(1)}, (Mc^2)^{(0)}\},$$

$$\tilde{U}^{(2)} = U^{(2)} + \{\lambda^{(2)}, (Mc^2)^{(0)}\} + \{\lambda^{(1)}, (Mc^2)^{(1)}\} \qquad (5.58)$$
$$+ \tfrac{1}{2}\{\lambda^{(1)}, \{\lambda^{(1)}, (Mc^2)^{(0)}\}\}, \ldots,$$

in agreement with the discussion given before, and $\chi_B^{(n)*}$, $\mathcal{X}_B^{(n)*}$ display a recurrent structure expressed only in terms of ρ, π, $\lambda^{(1)}, \lambda^{(2)}, \ldots, \lambda^{(n-1)}$ and $\chi_A^{(1)*}, \chi_A^{(2)*}, \ldots, \chi_A^{(n-1)*}$ or $\mathcal{X}_A^{(1)*}, \mathcal{X}_A^{(2)*}, \ldots, \mathcal{X}_A^{(n-1)*}$, respectively. For instance for $n = 2$ we have explicitly

$$\chi_B^{(2)*} = -\{\lambda^{(1)}, \chi_A^{(1)*}\} - \frac{1}{2}\left\{\lambda^{(1)}, \frac{\partial\lambda^{(1)}}{\partial\pi}\right\}$$

$$\mathcal{X}_B^{(2)*} = -\{\lambda^{(1)}, \mathcal{X}_A^{(1)*}\} - \left\{\lambda^{(1)}, \frac{\partial\Lambda^{(1)}}{\partial\pi}\right\}$$

$$- \left\{\lambda^{(1)}, \left(\frac{\pi_{20}}{\pi_{10} + \pi_{20}}\right)^{(1)} \rho\right\}. \qquad (5.59)$$

[Note for example, that the corresponding $\chi_1^{(2)*}(\rho, \pi)$ obtained in this way differs from the original choice made in Eq. (5.12) through the addition of the π gradient

$$\frac{\partial}{\partial\pi}\left[\tfrac{1}{2}\{\Lambda^{(1)}, \lambda^{(1)}\} + \frac{1}{m_1 m_2} \frac{\pi^2}{2} \Lambda^{(1)}\right].\Big] \qquad (5.60)]$$

It is then easy to show that the canonical transformation $T^{(k)} \cdot T^{(k-1)} \cdots T^{(2)} \cdot T^{(n)}$ eliminates just all of the $\chi_B^{(1)*}, \ldots, \chi_B^{(k)*}$, $\mathcal{X}_B^{(1)*}, \ldots, \mathcal{X}_B^{(k)*}$ with the consequence that, due to Eqs. (5.19), (5.19), the functions $\lambda^{(1)}, \ldots, \lambda^{(k)}$ are removed from the expression of $x_{1c.m.}$, $x_{2c.m.}$ and consequently of x_1, x_2. Finally, in force of the well-known Baker—Haussdorff formula[29] we can introduce a single canonical transformation

$$\tilde{T} = \exp[-\{\tilde{\lambda}, \cdots\}]$$
$$= \exp[-(1/c^2)\{\tilde{\lambda}^{(1)}, \cdots\} - (1/c^4)\{\tilde{\lambda}^{(2)}, \cdots\} + \cdots],$$
$$\qquad (5.61)$$

where $\tilde{\lambda}^{(n)}$ is derived from $\lambda^{(1)}, \lambda^{(2)}, \ldots, \lambda^{(3)}$ in a definite way. For example, up to the order $1/c^6$ we have

$$\tilde{\lambda}^{(1)} = \lambda^{(1)}, \quad \tilde{\lambda}^{(2)} = \lambda^{(2)},$$
$$\tilde{\lambda}^{(3)} = \lambda^{(3)} - \tfrac{1}{2}\{\lambda^{(1)}, \lambda^{(2)}\}. \qquad (5.62)$$

The transformation $\tilde{T}$ can then be interpreted as a global canonical equivalence acting on the whole realization of the Poincaré group. Any phase-space function $f(Q, P, \rho, \pi)$ maintains its functional form invariant under it apart from the replacement throughout of the original canonical potential $U(\rho, \pi)$ with its renormalized expression $\tilde{U}(\rho, \pi)$ and the elimination of the function λ. This is true in particular for the canonical generators of the group. We have

$$\tilde{P} = \tilde{T}^{-1}P = P, \quad \tilde{J} = \tilde{T}^{-1}J = J,$$
$$\tilde{H} = \tilde{T}^{-1}H = H[\tilde{U}], \quad \tilde{K} = \tilde{T}^{-1}K = K[\tilde{U}], \qquad (5.63)$$

with

$$\tilde{U} = \tilde{T}^{-1}U = U + [e^{\{\tilde{\lambda}, \cdots\}} - 1]Mc^2 \qquad (5.64)$$

or

$$\tilde{M}c^2 = \exp[\{\tilde{\lambda}, \cdots\}]Mc^2. \qquad (5.65)$$

We have exhausted so far the discussion of the formal arbitrariness left in the definition of the physical variables after the general conditions (a)—(d) of Sec. 3 and the "localizability" of the particle world lines have been taken into account. We have shown in conclusion that only two among the scalar functions $U(\rho, \pi)$, $\Lambda(\rho, \pi)$, $\lambda(\rho, \pi)$ can be physically significant. Returning to Eqs. (5.17) and putting $\lambda \equiv 0$, Eqs. (5.18) and (5.19) become

$$\mathbf{x}_{1_{\text{c.m.}}} - \mathbf{x}_{2_{\text{c.m.}}} = \rho + \chi^*[\rho, \pi, \tilde{U}, \Lambda], \quad (5.66)$$

$$\frac{\pi_{10}\mathbf{x}_{1_{\text{c.m.}}} + \pi_{20}\mathbf{x}_{2_{\text{c.m.}}}}{\pi_{10} + \pi_{20}}$$

$$= \mathbf{Q} + \chi^*[\rho, \pi, \tilde{U}, \Lambda] + \frac{\partial \Lambda}{\partial \pi}, \quad (5.67)$$

with

$$\Lambda(\rho, \pi) \equiv 0 \quad (\text{and } \chi^* \equiv 0), \quad (5.68)$$

in the particular case of *identical particles* $\kappa_1 \equiv \kappa_2$. Independently of the further possible restriction on the choice of Λ discussed above, we shall assume from now on that the "localizability" conditions have been enforced and we will consider the remaining arbitrariness with *emphasis on its physical meaning*. Actually, the dynamical significance of the interaction potentials $U(\rho, \pi)$, $\Lambda(\rho, \pi)$ and the "gauge" character of the function $\lambda(\rho, \pi)$ will appear in an expressive way throughout the whole discussion given in the following section. It is important, however, to realize since now that, under reasonable conditions, *the function* $\Lambda(\rho, \pi)$ [*and* $\lambda(\rho, \pi)$] *cannot affect the scattering behavior of the physical system*. In fact, let us consider a scattering experiment and denote by t' and t'' two instants of time far away the actual time of interaction, long before and after it, respectively. Since the potential $U(\rho, \pi)$ is the only dynamical element which enters the equations of motion of the basic canonical variables, the relations among $\mathbf{Q}(t''), \mathbf{P}(t''), \rho(t''), \pi(t'')$ and $\mathbf{Q}(t'), \mathbf{P}(t'), \rho(t'), \pi(t')$ can be affected only by $U(\rho, \pi)$. On the other hand if Eq. (3.18) is taken literally and thus the asymptotic condition (3.19) is assumed, the functions $U(\rho, \pi), \Lambda(\rho, \pi)$, $(\lambda(\rho, \pi))$ disappear from the expressions which give $\mathbf{x}_1(t'), \mathbf{x}_2(t'), \dot{\mathbf{x}}_1(t'), \dot{\mathbf{x}}_2(t')$ in terms of $\mathbf{Q}(t'), \mathbf{P}(t'), \rho(t')$, $\pi(t')$ [$\mathbf{x}_1(t''), \mathbf{x}_2(t''), \dot{\mathbf{x}}_1(t''), \dot{\mathbf{x}}_2(t'')$ in terms of $\mathbf{Q}(t'')$, $\mathbf{P}(t''), \rho(t''), \pi(t'')$]. This implies finally that the relation among $\mathbf{x}_1(t''), \mathbf{x}_2(t''), \dot{\mathbf{x}}_1(t''), \dot{\mathbf{x}}_2(t'')$, and $\mathbf{x}_1(t'), \mathbf{x}_2(t')$, $\dot{\mathbf{x}}_1(t'), \dot{\mathbf{x}}_2(t')$, which are the relevant ones for the description of the scattering, cannot depend on the function $\Lambda(\rho, \pi)$, $(\lambda(\rho, \pi))$ but are affected only by the canonical potential $U(\rho, \pi)$. The point is now whether Eq. (3.19) is consistent with the requirements subsequently made on the vectors $\chi_\tau(\rho, \pi)$, $(\tau = 1, 2)$, namely the "localizability" and possibly the "causality" conditions. To answer this point we may plainly assume that $\Lambda(\rho, \pi)$ [and $\lambda(\rho, \pi)$] behave asymptotically as $(\rho \cdot \pi) \cdot U(\rho, \pi)$ [see Eqs. (5.4), (5.44), and Sec. 6]. Then a direct check of Eqs. (5.4), (5.12) and the relevant equations of Sec. 6 shows that, provided $U(\rho, \pi)$ has short-range behavior, the asymptotic condition (3.19) is satisfied to the orders in $1/c^2$ which have been worked out explicitly and it is very likely to be satisfied to any order of the expansion. On the other hand if the asymptotic behavior of $U(\rho, \pi)$ is $O(1/\rho)$,

the same equations show that the vectors $\chi_\tau(\rho, \pi)$, $(\tau = 1, 2)$, are simply bounded for large ρ values. More precisely they are such that

$$\mathbf{x}_\tau \underset{\rho \to \infty}{\to} \mathbf{q}_\tau(\mathbf{Q}, \mathbf{P}, \rho, \pi) + O(1/\rho)\rho + O(1/\rho)(\rho \cdot \pi)\pi, \quad (\tau = 1, 2). \quad (5.69)$$

A careful analysis shows, however, that even in this case the relevant scattering parameters (the *impact parameter* and the *angles* defining the scattering direction) as functions of the basic canonical variables are not changed. An analogous situation occurs in connection with the possible bounded motions of the particles. Assuming internal "action" and "angle" variables to exist, *the dependence of the internal energy of the system on the action variables is affected again by the canonical potential* $U(\rho, \pi)$ *and not by the potential* $\Lambda(\rho, \pi)$ [*and the gauge* $\lambda(\rho, \pi)$]. Thus, in a corresponding quantized theory, the canonical potential U would be sufficient to construct the S matrix and evaluate bound state energies. [30] The "distinguishability potential" Λ is relevant only for the space—time description of the system in the region of the phase space where the interaction is important and thus it selects a particular theory within a class of theories having the same S-matrix equivalent. [31]

6. NEWTONIAN-LIKE EQUATIONS OF MOTION AND APPROXIMATELY-RELATIVISTIC LAGRANGIANS

In the preceding sections the canonical variables $\mathbf{Q}, \mathbf{P}, \rho, \pi$ were considered the *fundamental variables*. From now on we want to discuss the structure that the equations of motion and the whole canonical realization of the Poincaré group assume when expressed directly in terms of the physical variables $\mathbf{x}_1, \mathbf{x}_2, \dot{\mathbf{x}}_1, \dot{\mathbf{x}}_2$. The equations of motion can be obtained in principle by inverting the functions

$$\mathbf{x}_\tau = \mathbf{x}_\tau(\mathbf{Q}, \mathbf{P}, \rho, \pi), \quad (\tau = 1, 2), \quad (6.1)$$

and

$$\dot{\mathbf{x}}_\tau \equiv \{\mathbf{x}_\tau, H\} = \mathbf{v}_\tau(\mathbf{Q}, \mathbf{P}, \rho, \pi), \quad (\tau = 1, 2), \quad (6.2)$$

and introducing the resulting expressions into the rhs of the equations

$$\ddot{\mathbf{x}}_\tau = \{\{\mathbf{x}_\tau, H\}, H\}. \quad (6.3)$$

In this way it is possible to derive equations which share naturally a Newtonian-like form

$$m_\tau \ddot{\mathbf{x}}_\tau = \mathbf{F}_\tau[\mathbf{x}_1 - \mathbf{x}_2, \mathbf{v}_1, \mathbf{v}_2], \quad (\tau = 1, 2). \quad (6.4)$$

Alternatively, by introducing the proper times

$$ds_\tau = \sqrt{1 - \mathbf{v}_\tau^2/c^2} \, dt, \quad (\tau = 1, 2)$$

and the space part of the 4-velocity

$$\mathbf{u}_\tau = d\mathbf{x}_\tau/ds_\tau = \mathbf{v}_\tau/\sqrt{1 - \mathbf{v}_\tau^2/c^2}$$

the equations can be given the "covariant" form

$$m_\tau \frac{d^2\mathbf{x}_\tau}{ds_\tau^2} = \mathbf{f}_\tau[\mathbf{x}_1 - \mathbf{x}_2, \mathbf{u}_1, \mathbf{u}_2], \quad (\tau = 1, 2). \quad (6.5)$$

We shall presently give explicit expressions for the functions $\mathbf{F}_\tau$ and $\mathbf{f}_\tau$ in terms of the interaction potentials U, Λ and the gauge function λ under some significant as-

sumptions. To avoid complicated formulas we shall limit ourselves to discuss the terms of order $1/c^2$ of our expansion (Post-Newtonian approximation, PN). As we have recalled in the Introduction, this order of approximation is already highly significant in different contexts. What is more is that it will appear that *important features of the theory can manifest themselves only at precisely this order of relativistic approximation.*

We shall assume that:

(1) *The canonical interaction potential* $U(\rho,\pi)$ *possesses a static nonrelativistic limit*

$$U^{(0)}(\rho,\pi) = U^{(0)}_{[\kappa_1,\kappa_2]}(\rho);\tag{6.6}$$

(2) $U^{(1)}(\rho,\pi)$

has the structure

$$U^{(1)}(\rho,\pi) = \Xi_{[\kappa_1,\kappa_2]}(\rho) + \frac{\pi^2}{2\mu^2}\Phi_{[\kappa_1,\kappa_2]}(\rho) + \frac{\sigma^2}{2\mu^2}\Psi_{[\kappa_1,\kappa_2]}(\rho),\tag{6.7}$$

where the functions $\Xi_{[\kappa_1,\kappa_2]}(\rho)$, $\Phi_{[\kappa_1,\kappa_2]}(\rho)$, $\Psi_{[\kappa_1,\kappa_2]}(\rho)$ have asymptotic behavior and a symmetry in κ_1,κ_2 consistent with the conditions (b) and (c) of Sec. 2.

(3) The "localizability" conditions are satisfied up to any order of approximation while *the "causality" conditions are satisfied up to the PN approximation* $(1/c^2)$. This is consistent with Eq. (6.6) and implies moreover [see Eq. (5.37)]

$$\Lambda^{(1)}(\rho,\pi) = (1/2m)\Omega^{(1)}_{[\kappa_1,\kappa_2]}(\rho)\cdot\sigma.\tag{6.8}$$

(4) The PN gauge function $\lambda^{(1)}(\rho,\pi)$ has the structure

$$\lambda^{(1)}(\rho,\pi) = (1/m)\omega^{(1)}_{[\kappa_1,\kappa_2]}(\rho)\cdot\sigma.\tag{6.9}$$

[Of course we assume that also $\Omega^{(1)}(\rho)$ and $\omega^{(1)}(\rho)$ have all the required asymptotic properties.]
We emphasize that Eqs. (6.6), (6.7), (6.9) are all consequences of the general assumption that the *canonical potential* $U(\rho,\pi)$ and the *gauge function* $\lambda(\rho,\pi)$ have

the structure

$$U(\rho,\pi) = \gamma\hat{U}_{[\kappa_1,\kappa_2]}\left[\frac{\gamma}{\mu c^2}, \frac{\rho}{l}, \frac{\pi^2}{\mu^2 c^2}, \frac{\sigma^2}{l^2\mu^2 c^2}\right],\tag{6.10}$$

$$\lambda(\rho,\pi) = \frac{\gamma}{\mu c^2}\hat{\lambda}_{[\kappa_1,\kappa_2]}\left[\frac{\gamma}{\mu c^2}, \frac{\rho}{l}, \frac{\pi^2}{\mu^2 c^2}, \frac{\sigma^2}{l^2\mu^2 c^2}\right]\cdot\sigma,\tag{6.11}$$

with $\hat{U}_{[\kappa_1,\kappa_2]}$, $\hat{\lambda}_{[\kappa_1,\kappa_2]}$ dimensionless functions, i.e., a structure which contains the minimum possible number of dimensional constants besides the individual physical parameters of the particles and the light velocity, namely two constants γ and l with the dimensions of an energy and a length, respectively.

Note: While a single dimensional constant is suggested by interactions of gravitational or electromagnetic type, we consider two constants to describe also short-range forces, typically the forces generated by a massive field, which in the nonrelativistic limit clearly needs this number of dimensional constants, as the Yukawa potential. The long-range case in which a single constant appears is contained in our assumption as a particular case corresponding to a structure of the form

$$U(\rho,\pi) = \gamma\frac{l}{\rho}\hat{f}\left[\frac{\gamma l}{\mu c^2}\frac{1}{\rho}, \frac{1}{\mu^2 c^2}\pi^2, \frac{1}{\mu^2 c^2\rho^2}\sigma^2\right],$$

where γ and l are present only through the product γl. An example of this can be found in the Table at the end of the paper, case B.3, where $G = \gamma l/m_1 m_2$. Assuming the expansion (4.1) with $U^{(0)}(\rho,\pi)$ and $U^{(1)}(\rho,\pi)$ given by Eqs. (6.6) and (6.7), it is equivalent to assume that U is an analytic function of the zero-dimensional quantities $a = \gamma/\mu c^2$, $b = \pi^2/\mu^2 c^2$, $d = \sigma^2/\mu^2 c^2 l^2$ in a neighborhood of the origin. Then in particular we have

$$U^{(0)}(\rho) = \gamma[\hat{U}]_{(1/c^2=0)}, \quad \Xi(\rho) = \frac{\gamma^2}{\mu}\left[\frac{\partial\hat{U}}{\partial(\gamma/\mu c^2)}\right]_{(1/c^2=0)},$$

$$\Phi(\rho) = 2\mu\gamma\left[\frac{\partial\hat{U}}{\partial(\pi^2/\mu^2 c^2)}\right]_{(1/c^2=0)},$$

$$\Psi(\rho) = \frac{2\gamma}{l^2}\left[\frac{\partial\hat{U}}{\partial(\sigma^2/l^2\mu^2 c^2)}\right]_{(1/c^2=0)}.$$

Under the above assumptions, up to the PN approximation, we have

$$\mathbf{x}_1(\mathbf{Q},\mathbf{P},\rho,\pi) = \mathbf{x}_1^{(0)}(\mathbf{Q},\mathbf{P},\rho,\pi) + \frac{1}{c^2}\mathbf{x}_1^{(1)}(\mathbf{Q},\mathbf{P},\rho,\pi) + \cdots$$

$$= \mathbf{Q} + \frac{m_2}{m}\rho + \frac{1}{c^2}\left\{-\frac{m_2}{2m^3}(\rho\cdot\mathbf{P})\mathbf{P} + \left[\left(\frac{1}{m_2}-\frac{1}{m_1}\right)\frac{\pi^2}{2m} - \frac{\pi\cdot\mathbf{P}}{2m^2}\right]\rho + \frac{1}{2m^2}\left(1-\frac{2m_2}{m_1}\right)\cdot(\rho\cdot\mathbf{P})\pi\right.$$

$$\left. - \left(\frac{m_2}{m^2}U^{(0)}(\rho) - \frac{1}{2m}\Omega^{(1)}(\rho) - \frac{m_2}{m^2}\omega^{(1)}(\rho)\right)\rho\right\} + \cdots,\tag{6.12}$$

$$\mathbf{v}_1(\mathbf{Q},\mathbf{P},\rho,\pi) = \mathbf{v}_1^{(0)}(\mathbf{Q},\mathbf{P},\rho,\pi) + \frac{1}{c^2}\mathbf{v}_1^{(1)}(\mathbf{Q},\mathbf{P},\rho,\pi) + \cdots$$

$$= \frac{\mathbf{P}}{m} + \frac{\pi}{m_1} + \frac{1}{c^2}\left\{-\frac{1}{m^2}\left(H^{(0)} + \frac{\pi\cdot\mathbf{P}}{2m_1}\right)\mathbf{P} + \frac{1}{m}\left[\frac{m_2}{mm_1}(\rho\cdot\mathbf{P}) - \frac{1}{m_2}(\rho\cdot\pi)\right]\frac{1}{\rho}\frac{dU^{(0)}}{d\rho}\rho\right.$$

$$- \frac{1}{m}\left[H^{(0)} + \frac{m}{2m_1}\left(\frac{1}{m_1}-\frac{1}{m_2}\right)\pi^2 + \frac{1}{m_1}(\pi\cdot\mathbf{P})\right]\frac{\pi}{m_1} + \frac{m_2}{m}\left[\frac{1}{\mu^2}\Phi(\rho)\pi + \frac{1}{\mu^2}(\rho\cdot\pi)\Psi(\rho)\rho\right]$$

$$\left. + \frac{1}{2m_1 m_2}\left[(\rho\cdot\pi)\frac{1}{\rho}\frac{d\Omega^{(1)}}{d\rho}\rho + \Omega^{(1)}(\rho)\pi\right] + \frac{1}{mm_1}\left[(\rho\cdot\pi)\frac{1}{\rho}\frac{d\omega^{(1)}}{d\rho}\rho + \omega^{(1)}(\rho)\pi\right]\right\} + \cdots,\tag{6.13}$$

and the corresponding expressions for the particle 2 obtained by means of the $\mathcal{Z}$ operation (which implies in particular $\mathcal{Z}\omega^{(1)}(\rho) = \omega^{(1)}(\rho)$, $\mathcal{Z}\Omega^{(1)}(\rho) = -\Omega^{(1)}(\rho)$, $\mathcal{Z}U^{(0)}(\rho) = U^{(0)}(\rho)$, $\mathcal{Z}\Xi(\rho) = \Xi(\rho)$, $\mathcal{Z}\Phi(\rho) = \Phi(\rho)$, $\mathcal{Z}\Psi(\rho) = \Psi(\rho)$.

Then putting $\mathbf{r} \equiv \mathbf{x}_1 - \mathbf{x}_2$, we obtain by inversion

$$\mathbf{Q} = \frac{m_1 \mathbf{x}_1 + m_2 \mathbf{x}_2}{m} + \frac{1}{c^2}\left\{\frac{\mu}{2m^2}\left[(\mathbf{v}_1 - \mathbf{v}_2)\cdot(m_1\mathbf{v}_1 + m_2\mathbf{v}_2) - (m_1 - m_2)(\mathbf{v}_1 - \mathbf{v}_2)^2\right]\mathbf{r} + \frac{\mu}{2m^2}[\mathbf{r}\cdot(m_1\mathbf{v}_1 + m_2\mathbf{v}_2)](\mathbf{v}_1 - \mathbf{v}_2)\right.$$

$$\left. - \frac{1}{2m}\Omega^{(1)}(r)\mathbf{r}\right\} + \cdots,$$

$$\mathbf{P} = m_1\mathbf{v}_1 + m_2\mathbf{v}_2 + \frac{1}{c^2}\left\{\frac{1}{m}\left[\tfrac{1}{2}m_1\mathbf{v}_1^2 + \tfrac{1}{2}m_2\mathbf{v}_2^2 + U^{(0)}(r)\right](m_1\mathbf{v}_1 + m_2\mathbf{v}_2) + \frac{1}{m}\left[(m_1 - m_2)\mathbf{r}\cdot(\mathbf{v}_1 - \mathbf{v}_2) - \mathbf{r}\cdot(m_1\mathbf{v}_1 + m_2\mathbf{v}_2)\right]\frac{1}{r}\frac{dU^{(0)}}{dr}\mathbf{r}\right.$$

$$\left. + \frac{1}{m}\left[\mu(\mathbf{v}_1 - \mathbf{v}_2)\cdot(m_1\mathbf{v}_1 + m_2\mathbf{v}_2) - (m_1 - m_2)\tfrac{1}{2}\mu(\mathbf{v}_1 - \mathbf{v}_2)^2\right](\mathbf{v}_1 - \mathbf{v}_2) - \tfrac{1}{2}\Omega^{(1)}(r)(\mathbf{v}_1 - \mathbf{v}_2) - \tfrac{1}{2}\mathbf{r}\cdot(\mathbf{v}_1 - \mathbf{v}_2)\frac{1}{r}\frac{d\Omega^{(1)}}{dr}\mathbf{r}\right\} + \cdots,$$

$$\boldsymbol{\rho} = \mathbf{r} + \frac{1}{c^2}\left\{\frac{1}{m}U^{(0)}(r)\mathbf{r} - \left[\frac{m_1 - m_2}{m^2}\mathbf{r}\cdot(m_1\mathbf{v}_1 + m_2\mathbf{v}_2)\right](\mathbf{v}_1 - \mathbf{v}_2) + \left[\frac{1}{2m^2}\mathbf{r}\cdot(m_1\mathbf{v}_1 + m_2\mathbf{v}_2)\right](m_1\mathbf{v}_1 + m_2\mathbf{v}_2)\right.$$

$$\left. - \frac{1}{m}\omega^{(1)}(r)\mathbf{r}\right\} + \cdots,$$

(6.14)

$$\boldsymbol{\pi} = \mu(\mathbf{v}_1 - \mathbf{v}_2) + \frac{1}{c^2}\left\{\left[\frac{\mu}{2m^2}(\mathbf{v}_1 - \mathbf{v}_2)(m_1\mathbf{v}_1 + m_2\mathbf{v}_2)\right](m_1\mathbf{v}_1 + m_2\mathbf{v}_2) + \left[\frac{m_1 - m_2}{m^2}\mathbf{r}\cdot(m_1\mathbf{v}_1 + m_2\mathbf{v}_2)\right.\right.$$

$$\left. + \frac{\mu}{m}\mathbf{r}\cdot(\mathbf{v}_1 - \mathbf{v}_2)\right]\frac{1}{r}\frac{dU^{(0)}}{dr}\mathbf{r} + \left[\frac{\mu}{m}\left[\tfrac{1}{2}m_1\mathbf{v}_1^2 + \tfrac{1}{2}m_2\mathbf{v}_2^2 + U^{(0)}(r)\right] + (m_1 - m_2)\frac{\mu}{2m}(\mathbf{v}_1 - \mathbf{v}_2)^2\right.$$

$$\left. - (m_1 - m_2)\frac{\mu}{m^2}(\mathbf{v}_1 - \mathbf{v}_2)\cdot(m_1\mathbf{v}_1 + m_2\mathbf{v}_2)\right](\mathbf{v}_1 - \mathbf{v}_2) - \frac{1}{\mu}\Phi(r)(\mathbf{v}_1 - \mathbf{v}_2) - \frac{1}{\mu}\mathbf{r}\cdot(\mathbf{v}_1 - \mathbf{v}_2)\Psi(r)\mathbf{r}$$

$$\left. - \frac{\mu}{m}\omega^{(1)}(r)(\mathbf{v}_1 - \mathbf{v}_2) - \frac{\mu}{m}\mathbf{r}\cdot(\mathbf{v}_1 - \mathbf{v}_2)\frac{1}{r}\frac{d\omega^{(1)}}{dr}\mathbf{r}\right\} + \cdots.$$

From Eqs. (6.3), (6.4), (6.14) it follows that

$$\mathbf{F}_1 = \mathbf{F}_1^{(0)} + \frac{1}{c^2}\mathbf{F}_1^{(1)} + \cdots$$

$$= -\frac{1}{r}\frac{dU^{(0)}}{dr}\mathbf{r} + \frac{1}{c^2}\left\{\left\langle\left[\frac{1}{m}(\tfrac{1}{2}m_1\mathbf{v}_1^2 + \tfrac{1}{2}m_2\mathbf{v}_2^2) + \frac{1}{2m^2}(m_1\mathbf{v}_1 + m_2\mathbf{v}_2)^2 + \frac{2m_2}{m^2}(\mathbf{v}_1 - \mathbf{v}_2)\cdot(m_1\mathbf{v}_1 + m_2\mathbf{v}_2)\right.\right.\right.$$

$$\left. + \frac{m_2^2 - m_1 m_2 - 2m_1^2}{2m^2}(\mathbf{v}_1 - \mathbf{v}_2)^2 - \frac{m_1}{\mu}\frac{1}{2m}\Omega^{(1)}(r) - \frac{m_1}{\mu}\frac{1}{2m}r\frac{d\Omega^{(1)}}{dr} - \frac{1}{\mu}\Phi(r) - \frac{1}{\mu}r^2\Psi(r) - \frac{1}{m}r\frac{d\omega^{(1)}}{dr}\right]\frac{1}{r}\frac{dU^{(0)}}{dr}$$

$$+ \left[\frac{m_1}{m^2}\mathbf{r}\cdot(m_1\mathbf{v}_1 + m_2\mathbf{v}_2)\mathbf{r}\cdot(\mathbf{v}_1 - \mathbf{v}_2) - \frac{1}{2m^2}(\mathbf{r}\cdot(m_1\mathbf{v}_1 + m_2\mathbf{v}_2))^2 - \frac{m_1^2}{m^2}(\mathbf{r}\cdot(\mathbf{v}_1 - \mathbf{v}_2))^2 - \frac{1}{m}U^{(0)}(r)r^2\right.$$

$$+ \frac{1}{m}r^2\omega^{(1)}(r)\right]\left(\frac{1}{r}\frac{d}{dr}\right)^2 U^{(0)}(r) + \left[\frac{m_1}{mm_2}\left(\frac{dU^{(0)}}{dr}\right)^2\right] + \left[\frac{m_1}{2m}(\mathbf{v}_1 - \mathbf{v}_2)^2\frac{1}{r}\frac{d\Omega^{(1)}}{dr} + \frac{m_1}{2m}(\mathbf{r}\cdot(\mathbf{v}_1 - \mathbf{v}_2))^2\left(\frac{1}{r}\frac{d}{dr}\right)^2\Omega^{(1)}(r)\right.$$

$$+ (\mathbf{v}_1 - \mathbf{v}_2)^2\Psi(r) + \tfrac{1}{2}(\mathbf{r}\cdot(\mathbf{v}_1 - \mathbf{v}_2))^2\frac{1}{r}\frac{d\Psi(r)}{dr} - \tfrac{1}{2}(\mathbf{v}_1 - \mathbf{v}_2)^2\frac{1}{r}\frac{d\Phi(r)}{dr} - \frac{1}{r}\frac{d\Xi(r)}{dr} + \frac{\mu}{m}(\mathbf{v}_1 - \mathbf{v}_2)^2\frac{1}{r}\frac{d\omega^{(1)}}{dr}$$

$$\left. + \frac{\mu}{m}(\mathbf{r}\cdot(\mathbf{v}_1 - \mathbf{v}_2))^2\left(\frac{1}{r}\frac{d}{dr}\right)^2\omega^{(1)}(r)\right]\right\rangle\mathbf{r} + \left\langle\left[\frac{1}{m}\mathbf{r}\cdot(m_1\mathbf{v}_1 + m_2\mathbf{v}_2) + \frac{1}{m^2}(m_2^2 - m_1^2 - m_1 m_2)\mathbf{r}\cdot(\mathbf{v}_1 - \mathbf{v}_2)\right]\frac{1}{r}\frac{dU^{(0)}}{dr}\right.$$

$$\left. + \left[\frac{1}{m^2}\mathbf{r}\cdot(\mathbf{v}_1 - \mathbf{v}_2)\frac{1}{r}\frac{d\Omega^{(1)}}{dr} + \frac{2\mu}{m}\mathbf{r}\cdot(\mathbf{v}_1 - \mathbf{v}_2)\frac{1}{r}\frac{d\omega^{(1)}}{dr} + \mathbf{r}\cdot(\mathbf{v}_1 - \mathbf{v}_2)\frac{1}{r}\frac{d\Phi(r)}{dr}\right]\right\rangle(\mathbf{v}_1 - \mathbf{v}_2)\right\} + \cdots,$$

(6.15)

or, from Eqs. (6.5), the expressions

$$\mathbf{f}_1 = \mathbf{f}_1^{(0)} + \frac{1}{c^2}\mathbf{f}_1^{(1)} + \cdots$$

$$= -\frac{1}{r}\frac{dU^{(0)}}{dr}\mathbf{r} + \frac{1}{c^2}\left\{\left\langle\left[\frac{r^2}{m}(\omega^{(1)}(r) - U^{(0)}(r))\left(\frac{1}{r}\frac{d}{dr}\right)^2 U^{(0)}(r) + \left(\frac{m_1}{mm_2}r\frac{dU^{(0)}}{dr} - \frac{1}{m}r\frac{d\omega^{(1)}}{dr} - \frac{1}{2m_2}\Omega^{(1)}(r)\right.\right.\right.$$

$$\left. - \frac{1}{2m_2}r\frac{d\Omega^{(1)}}{dr} - \frac{1}{\mu}\Phi(r) - \frac{r^2}{\mu}\Psi(r)\right)\frac{1}{r}\frac{dU^{(0)}}{dr} - \frac{1}{r}\frac{d\Xi(r)}{dr}\right] + (\mathbf{u}_1 - \mathbf{u}_2)^2\left[-\frac{1}{2}\left(1 + \frac{m_1^2}{m^2}\left(1 - \frac{2m_2}{m_1}\right)\right)\frac{1}{r}\frac{dU^{(0)}}{dr}\right.$$

$$\left. + \frac{m_1}{2m}\frac{1}{r}\frac{d\Omega^{(1)}}{dr} + \frac{\mu}{m}\frac{1}{r}\frac{d\omega^{(1)}}{dr} - \frac{1}{2}\frac{1}{r}\frac{d\Phi(r)}{dr} + \Psi(r)\right] + (\mathbf{r}\cdot\mathbf{u}_2)^2\left[-\frac{1}{2}\cdot\left(\frac{1}{r}\frac{d}{dr}\right)^2 U^{(0)}(r)\right]$$

$$\left. + ((\mathbf{r}\cdot\mathbf{u}_1) - (\mathbf{r}\cdot\mathbf{u}_2))^2\left[-\frac{m_1^2}{2m^2}\left(\frac{1}{r}\frac{d}{dr}\right)^2 U^{(0)}(r) + \frac{m_1}{2m}\left(\frac{1}{r}\frac{d}{dr}\right)^2\Omega^{(1)}(r) + \frac{1}{2}\frac{1}{r}\frac{d\Psi(r)}{dr} + \frac{\mu}{m}\left(\frac{1}{r}\frac{d}{dr}\right)^2\omega^{(1)}(r)\right]\right\rangle\mathbf{r}$$

$$+ \left\langle (\mathbf{r} \cdot \mathbf{u}_1) \left(-\frac{1}{r} \frac{dU^{(0)}}{dr} \right) \right\rangle \mathbf{u}_2 + \left\{ ((\mathbf{r} \cdot \mathbf{u}_1) - (\mathbf{r} \cdot \mathbf{u}_2)) \left\langle \frac{m_1}{m} \frac{1}{r} \frac{d\Omega^{(1)}}{dr} + \frac{1}{r} \frac{d\Phi(r)}{dr} + \frac{2\mu}{m} \frac{1}{r} \frac{d\omega^{(1)}}{dr} \right. \right.$$

$$\left. \left. - \frac{m_1}{m} \left(1 + \frac{m_2}{m} \right) \frac{1}{r} \frac{dU^{(0)}}{dr} \right] \right\rangle (\mathbf{u}_1 - \mathbf{u}_2) \right\} + \cdots . \tag{6.16}$$

The corresponding expressions for $\mathbf{F}_2$ and $\mathbf{f}_2$ can be obtained by means of the Z operation which implies in particular $m_1 \rightleftarrows m_2$, $\mathbf{r} \rightarrow -\mathbf{r}$ and $\mathbf{v}_1 \rightleftarrows \mathbf{v}_2$ or $\mathbf{u}_1 \rightleftarrows \mathbf{u}_2$, respectively.

A better insight into the structure of the functions $\mathbf{f}_1^{(0)}, \mathbf{f}_1^{(1)}, \cdots$, as given by Eq. (6.16) can be achieved if we introduce the 4-forces

$$f_\tau^\mu \equiv [f_\tau^0, \mathbf{f}_\tau], \quad (\tau = 1, 2), \tag{6.17}$$

where f_τ^0 is defined by

$$u_\tau^0 f_\tau^0 = \mathbf{u}_\tau \cdot \mathbf{f}_\tau, \quad (\tau = 1, 2), \tag{6.18}$$

and rewrite Eqs. (6.5) in the four-dimensional form

$$m_\tau \frac{d^2 x_\tau^\mu}{ds_\tau^2} = f_\tau^\mu (\mathbf{x}_1, \mathbf{x}_2, \mathbf{u}_1, \mathbf{u}_2), \quad (\tau = 1, 2). \tag{6.19}$$

As a matter of fact, in force of relativistic invariance, we must have

$$f_\tau^\mu = \varphi_\tau r^\mu + \psi_{\tau 1} u_1^\mu + \psi_{\tau 2} u_2^\mu, \quad (\tau = 1, 2), \tag{6.20}$$

where $r^\mu = x_1^\mu - x_2^\mu$ and $\varphi_\tau, \psi_{\tau 1}, \psi_{\tau 2}$ $(\tau = 1, 2)$ are relativistic scalars and thus functions of the independent invariants

$$\xi^2 \equiv -r_\mu r^\mu = r^2, \quad \eta_1 \equiv -r_\mu u_1^\mu = \mathbf{r} \cdot \mathbf{u}_1, \quad \eta_2 \equiv -r_\mu u_2^\mu = \mathbf{r} \cdot \mathbf{u}_2,$$

$$\theta \equiv -(u_1 - u_2)_\mu (u_1 - u_2)^\mu = (\mathbf{u}_1 - \mathbf{u}_2)^2 - (1/c^2)\tfrac{1}{4}(u_1^2 - u_2^2)^2 + O(1/c^4), \tag{6.21}$$

where the rhs expressions are consequences of the fact that the points x_1 and x_2 (or s_1, s_2) on the world lines *are assumed to be simultaneous in the considered reference frame*. Owing to Eq. (6.18) it follows in particular

$$\psi_{11} = -\psi_{12} + \frac{1}{c^2} (\eta_1 \varphi_1 - \tfrac{1}{2}\theta \psi_{12}), \quad \psi_{22} = -\psi_{21} + \frac{1}{c^2} (\eta_1 \varphi_2 - \tfrac{1}{2}\theta \psi_{21}), \tag{6.22}$$

and consequently

$$f_1^\mu = \varphi_1 r^\mu - \psi_{12}(u_1^\mu - u_2^\mu) + \frac{1}{c^2} (\eta_1 \varphi_1 - \tfrac{1}{2}\theta \psi_{12}) u_1^\mu,$$

$$f_2^\mu = \varphi_2 r^\mu + \psi_{21}(u_1^\mu - u_2^\mu) + \frac{1}{c^2} (\eta_2 \varphi_2 - \tfrac{1}{2}\theta \psi_{21}) u_2^\mu. \tag{6.23}$$

Then, for $\tau = 1$, by expanding φ_1 and ψ_{12}, we obtain for the space-part of the 4-force in the PN approximation

$$f_1^{(0)} = \varphi_1^{(0)} \mathbf{r} - \psi_{12}^{(0)}(\mathbf{u}_1 - \mathbf{u}_2),$$

$$f_1^{(1)} = \varphi_1^{(1)} \mathbf{r} + (\eta_1 \varphi_1^{(0)} - \tfrac{1}{2}\theta \psi_{12}^{(0)})\mathbf{u}_2 + (\eta_1 \varphi_1^{(0)} - \tfrac{1}{2}\theta \psi_{12}^{(0)} - \psi_{12}^{(1)})(\mathbf{u}_1 - \mathbf{u}_2). \tag{6.24}$$

Finally, taking into account Eqs. (6.21) it easily seen that Eqs. (6.16) agree with Eqs. (6.24) if

$$\psi_{12}^{(0)} \equiv 0, \tag{6.25}$$

which is a consequence of the static nonrelativistic limit assumption, and

$$\varphi_1^{(0)}(\xi, \eta_1, \eta_2, \theta) = -\frac{1}{\xi} \frac{dU^{(0)}(\xi)}{d\xi},$$

$$\varphi_1^{(1)}(\xi, \eta_1, \eta_2, \theta) = \left[\frac{1}{m} \xi^2 (\omega^{(1)}(\xi) - U^{(0)}(\xi)) \left(\frac{1}{\xi} \frac{d}{d\xi} \right)^2 U^{(0)}(\xi) + \left(\frac{m_1}{mm_2} \xi \frac{dU^{(0)}}{d\xi} - \frac{1}{m} \xi \frac{d\omega^{(1)}(\xi)}{d\xi} - \frac{1}{2m_2} \Omega^{(1)}(\xi) \right. \right.$$

$$- \frac{1}{2m_2} \xi \frac{d\Omega^{(1)}(\xi)}{d\xi} - \frac{1}{\mu} \Phi(\xi) - \frac{\xi^2}{\mu} \Psi(\xi) \left. \right) \frac{1}{\xi} \frac{dU^{(0)}(\xi)}{d\xi} - \frac{1}{\xi} \frac{d\Xi(\xi)}{d\xi} \right] + \tfrac{1}{2}\theta \left[-\left(1 + \frac{m_1^2}{m^2} \left(1 - \frac{2m_2}{m_1} \right) \right) \frac{1}{\xi} \frac{dU^{(0)}}{d\xi} \right.$$

$$+ \frac{m_1}{m} \frac{1}{\xi} \frac{d\Omega^{(1)}(\xi)}{d\xi} + \frac{2\mu}{m} \frac{1}{\xi} \frac{d\omega^{(1)}(\xi)}{d\xi} - \frac{1}{\xi} \frac{d\Phi(\xi)}{d\xi} + 2\Psi(\xi) \right] + \eta_2^2 \left[-\frac{1}{2} \left(\frac{1}{\xi} \frac{d}{d\xi} \right)^2 U^{(0)}(\xi) \right] \tag{6.26}$$

$$+ (\eta_1 - \eta_2)^2 \left[-\frac{m_1^2}{2m^2} \left(\frac{1}{\xi} \frac{d}{d\xi} \right)^2 U^{(0)}(\xi) + \frac{\mu}{m} \left(\frac{1}{\xi} \frac{d}{d\xi} \right)^2 \omega^{(1)}(\xi) + \frac{m_1}{2m} \left(\frac{1}{\xi} \frac{d}{d\xi} \right)^2 \Omega^{(1)}(\xi) + \frac{1}{2} \frac{1}{\xi} \frac{d\Psi(\xi)}{d\xi} \right]$$

$$\psi_{12}^{(1)}(\xi, \eta_1, \eta_2, \theta) = -\eta_1 \left[\frac{1}{\xi} \frac{dU^{(0)}(\xi)}{d\xi} \right] + (\eta_2 - \eta_1) \left[\frac{m_1}{m} \frac{1}{\xi} \frac{d\Omega^{(1)}(\xi)}{d\xi} + \frac{1}{\xi} \frac{d\Phi(\xi)}{d\xi} \right.$$

$$+ \frac{2\mu}{m} \frac{1}{\xi} \frac{d\omega^{(1)}(\xi)}{d\xi} - \frac{m_1}{m} \left(1 + \frac{m_2}{m} \right) \frac{1}{\xi} \frac{dU^{(0)}(\xi)}{d\xi} \right].$$

It is hardly necessary to emphasize that the dynamical variables of the two particles in Eqs. (6.19)—(6.26) must be evaluated at the same time with respect to the considered reference frame [recall Eqs. (6.21)!]. It is also possible, however, in force of the intrinsic co-variant character of Eqs. (6.19), to obtain a system of equations of motion in terms of the dynamical variables evaluated at two points on the world lines of the particles which share an *arbitrary spacelike relation*. This system is simply of the form

$$m_\tau \frac{d^2 x_\tau^\mu}{ds_\tau^2} = S_{.\nu}^\mu f_\tau^\nu [x_1(s_1), x_2(s_2), u_1(s_1), u_2(s_2)], \qquad (6.27)$$

where s_1 and s_2 are the proper times which specify the positions of the two points on the world lines and $S_{.\nu}^\mu$ is the Lorentz transformation which connects the considered reference frame with a second one in which the two points are simultaneous. In this way we have at our disposal *infinitely many descriptions in terms of manifestly covariant equations of motion corresponding to the same physical theory*. If we consider the sums of all the series expansions assuming analyticity in $1/c^2$, we find that Eqs. (6.27) are simply related to the manifestly covariant equations proposed by Havas and Plebański.[15] These equations correspond to a choice of different $S_{.\nu}^\mu$ for the two particles and thus to a different space—time relation between $x_1(s_1)$ and $x_2(s_2)$ for the equations of motion of particle 1 and particle 2; precisely, the prescriptions given by Havas and Plebański are

$$\eta_1 \equiv [x_1^\mu(s_1) - x_2^\mu(s_2)] u_{1\mu}(s_1) = 0,$$

for the equation of motion of particle 1,

$$\eta_2 \equiv [x_1^\mu(s_1) - x_2^\mu(s_2)] u_{2\mu}(s_2) = 0 \qquad (6.28)$$

for the equation of motion of particle 2.

Note however that the Havas—Plebański equations are not, strictly speaking, a system of closed equations.

We want to discuss now the interesting problem of the construction of a Lagrangian function from which our Newtonian equations of motion in the form (6.4) can be derived up to a certain order of approximation. In this connection it is important to realize first that it is certainly possible to derive Eqs. (6.4) up to any order from a variational principle of the form

$$\delta \int \left[\sum_{\tau=1}^2 (\dot{\mathbf{x}}_\tau \cdot \mathcal{P}_\tau(\mathbf{x}_1, \mathbf{x}_2, \mathbf{v}_1, \mathbf{v}_2) + \dot{\mathbf{v}}_\tau \cdot \mathcal{R}_\tau(\mathbf{x}_1, \mathbf{x}_2, \mathbf{v}_1, \mathbf{v}_2)) \right.$$
$$\left. - \mathcal{W}(\mathbf{x}_1, \mathbf{x}_2, \mathbf{v}_1, \mathbf{v}_2) \right] dt = 0 \qquad (6.29)$$

(see for example Refs. 19 and 20). (We do not consider here the possibility of constructing "Lagrangians" containing infinite series involving v_τ and derivatives.) Actually the variational principle (6.29) can be immediately obtained from the obvious variational principle in the canonical form

$$\delta \int [\dot{\mathbf{Q}} \cdot \mathbf{P} + \dot{\boldsymbol{\rho}} \cdot \boldsymbol{\pi} - H(\mathbf{Q}, \mathbf{P}, \boldsymbol{\rho}, \boldsymbol{\pi})] dt = 0, \qquad (6.30)$$

by re-expressing the variables $\mathbf{Q}, \mathbf{P}, \boldsymbol{\rho}, \boldsymbol{\pi}$ in terms of

$\mathbf{x}_1, \mathbf{x}_2, \mathbf{v}_1, \mathbf{v}_2$. It follows that

$$\mathcal{P}_{\tau i} = \mathbf{P} \cdot \frac{\partial \mathbf{Q}}{\partial x_{\tau i}} + \boldsymbol{\pi} \cdot \frac{\partial \boldsymbol{\rho}}{\partial x_{\tau i}},$$

$$\mathcal{R}_{\tau j} = \mathbf{P} \cdot \frac{\partial \mathbf{Q}}{\partial v_{\tau j}} + \boldsymbol{\pi} \cdot \frac{\partial \boldsymbol{\rho}}{\partial v_{\tau j}}, \qquad (6.31)$$

$$\mathcal{W} = H.$$

On the contrary, *it is not possible to derive the same equations from a true Lagrangian*, i.e., from a variational principle of the form

$$\delta \int L(\mathbf{x}_1, \mathbf{x}_2, \dot{\mathbf{x}}_1, \dot{\mathbf{x}}_2) \, dt = 0. \qquad (6.32)$$

Of course the essential difference between (6.29) and (6.32) lies in the fact that while the *true Lagrangian* L depends only on the variables $\mathbf{x}_1, \mathbf{x}_2, \dot{\mathbf{x}}_1, \dot{\mathbf{x}}_2$, the integrand appearing in Eq. (5.29) depends in addition on the variables $\dot{\mathbf{v}}_1 = \ddot{\mathbf{x}}_1, \dot{\mathbf{v}}_2 = \ddot{\mathbf{x}}_2$. As a matter of fact, as noted by Kerner,[12] the existence of *a variational principle of the true Lagrangian form* would allow a Hamiltonization of the Newtonian-like equations of motion along the usual route given by the Legendre transformation. This in turn, via the Noether theorem, would also allow the construction of a canonical realization of the Poincaré group in which the physical position vectors $\mathbf{x}_1$, $\mathbf{x}_2$ would play the role of canonical (configurational) variables, a fact which would contradict the zero-interaction theorem.

The zero-interaction theorem, however, cannot prevent the "localizability" and "causality" conditions from being satisfied *up to a definite order in* $1/c^2$ as it is clear from the discussion of the preceding section. This allows in fact the existence of an *approximately—relativistic* Lagrangian in the following precise sense. Taking into account the expressions (6.14), Eqs. (6.31) become (for $\tau = 1$)

$$\mathcal{P}_{1i} = m_1 v_{1i} + \frac{1}{c^2} \left(\mathbf{P}^{(0)} \cdot \frac{\partial \mathbf{Q}^{(1)}}{\partial x_{1i}} + \boldsymbol{\pi}^{(0)} \cdot \frac{\partial \boldsymbol{\rho}^{(1)}}{\partial x_{1i}} \right.$$
$$\left. + \mathbf{P}^{(1)} \cdot \frac{\partial \mathbf{Q}^{(0)}}{\partial x_{1i}} + \boldsymbol{\pi}^{(1)} \cdot \frac{\partial \boldsymbol{\rho}^{(0)}}{\partial x_{1i}} \right) + O\left(\frac{1}{c^4}\right),$$

$$\mathcal{R}_{1j} = \frac{1}{c^2} \left(\mathbf{P}^{(0)} \cdot \frac{\partial \mathbf{Q}^{(1)}}{\partial v_{1j}} + \boldsymbol{\pi}^{(0)} \cdot \frac{\partial \boldsymbol{\rho}^{(1)}}{\partial v_{1j}} \right) + O\left(\frac{1}{c^4}\right) \qquad (6.33)$$

$$= \frac{1}{c^2} \frac{\partial}{\partial v_{1j}} \{ [m_2 \mathbf{Q}^{(1)} - \mu \boldsymbol{\rho}^{(1)}] \cdot \mathbf{v}_2 \} + O\left(\frac{1}{c^4}\right),$$

and analogous expressions for $\mathcal{P}_{2i}$ and $\mathcal{R}_{2j}$.

On the other hand the fact that the integrand of Eq. (6.29) is obviously determined up to a total derivative implies that the expressions $\mathcal{P}_\tau$, $\mathcal{R}_\tau$, and $\mathcal{W}$ are determined up to a transformation of the form

$$\mathcal{P}_\tau \to \mathcal{P}'_\tau = \mathcal{P}_\tau - \frac{\partial \Theta}{\partial \mathbf{x}_\tau},$$

$$\mathcal{R}_\tau \to \mathcal{R}'_\tau = \mathcal{R}_\tau - \frac{\partial \Theta}{\partial \mathbf{v}_\tau}, \qquad (6.34)$$

$$\mathcal{W} \to \mathcal{W}' = \mathcal{W} + \frac{\partial \Theta}{\partial t}.$$

Therefore choosing

$$\Theta = \frac{1}{c^2} \{ [m_2 \mathbf{Q}^{(1)}(\mathbf{r}, \mathbf{v}_1, \mathbf{v}_2) - \mu \boldsymbol{\rho}^{(1)}(\mathbf{r}, \mathbf{v}_1, \mathbf{v}_2)] \cdot \mathbf{v}_2 \}, \qquad (6.35)$$

we obtain

$$p_1' = m_1\mathbf{v}_1 + \frac{1}{c^2}\frac{1}{2}\left\{ m_1\mathbf{v}_1^2\mathbf{v}_1 + \left[\frac{2m_1}{m^2}(m_1\mathbf{v}_1 + m_2\mathbf{v}_2)\right.\right.$$
$$+ \frac{4\mu}{m}(\mathbf{v}_1 - \mathbf{v}_2)\bigg]U^{(0)}(r) + \left[\frac{2m_1}{m^2}\mathbf{r}\cdot(m_1\mathbf{v}_1 + m_2\mathbf{v}_2)\right.$$
$$- 2(\mathbf{r}\cdot\mathbf{v}_2)\bigg]\frac{1}{r}\frac{dU^{(0)}}{dr}\,\mathbf{r} + \left[\frac{1}{m}((m_1 - m_2)\mathbf{v}_2 - 2m_1\mathbf{v}_1)\right]\Omega^{(1)}(r)$$
$$+ \left[\frac{1}{m}((m_1 - m_2)(\mathbf{r}\cdot\mathbf{v}_2) - 2m_1\cdot(\mathbf{r}\cdot\mathbf{v}_1)\right]\frac{1}{r}\cdot\frac{d\Omega^{(1)}}{dr}$$
$$\times \mathbf{r} - \frac{4\mu}{m}\omega^{(1)}(r)(\mathbf{v}_1 - \mathbf{v}_2) - \frac{4\mu}{m}\mathbf{r}\cdot(\mathbf{v}_1 - \mathbf{v}_2)\frac{1}{r}\frac{d\omega^{(1)}}{dr}\mathbf{r}$$
$$- 2\Phi(r)(\mathbf{v}_1 - \mathbf{v}_2) - 2\mathbf{r}\cdot(\mathbf{v}_1 - \mathbf{v}_2)\,\Psi(r)\mathbf{r}\bigg\} + O\left(\frac{1}{c^4}\right)$$

$$p_2' = m_2\mathbf{v}_2 + \frac{1}{c^2}\frac{1}{2}\left\{ m_2\mathbf{v}_2^2\mathbf{v}_2 + \left[\frac{2m_2}{m^2}(m_1\mathbf{v}_1 + m_2\mathbf{v}_2)\right.\right.$$
$$- \frac{4\mu}{m}(\mathbf{v}_1 - \mathbf{v}_2)\bigg]U^{(0)}(r) + \left[\frac{2m_2}{m^2}\mathbf{r}\cdot(m_1\mathbf{v}_1 + m_2\mathbf{v}_2)\right.$$
$$- 2(\mathbf{r}\cdot\mathbf{v}_1)\bigg]\frac{1}{r}\frac{dU^{(0)}}{dr}\mathbf{r} + \left[\frac{1}{m}((m_1 - m_2)\mathbf{v}_1\right.$$
$$+ 2m_2\mathbf{v}_2\bigg]\Omega^{(1)}(r) + \left[\frac{1}{m}((m_1 - m_2)(\mathbf{r}\cdot\mathbf{v}_1)\right.$$
$$+ 2m_2(\mathbf{r}\cdot\mathbf{v}_2))\bigg]\frac{1}{r}\frac{d\Omega^{(1)}}{dr}\mathbf{r} + \frac{4\mu}{m}\omega^{(1)}(r)(\mathbf{v}_1 - \mathbf{v}_2)$$
$$+ \frac{4\mu}{m}\mathbf{r}\cdot(\mathbf{v}_1 - \mathbf{v}_2)\frac{1}{r}\frac{d\omega^{(1)}}{dr}\mathbf{r} + 2\Phi(r)(\mathbf{v}_1 - \mathbf{v}_2)$$
$$+ 2\mathbf{r}\cdot(\mathbf{v}_1 - \mathbf{v}_2)\Psi(r)\mathbf{r}\bigg\} + O\left(\frac{1}{c^4}\right),$$

(6.36)

$$R_1' = O(1/c^4),$$

$$R_2' = O(1/c^4),$$

$$W' = W = H.$$

We see in this way that the variational principle (6.29) *takes a true Lagrangian form up to the PN approximation. Stated in other words, the first two terms* $\mathbf{F}_\tau^{(0)}$, ($\tau = 1, 2$), *and* $\mathbf{F}_\tau^{(1)}$, ($\tau = 1, 2$) *of the Newtonian-like equations of motion* (6.15) *can be derived from the approximately relativistic Lagrangian*

$$L = [p_1'\cdot\mathbf{v}_1 + p_2'\cdot\mathbf{v}_2 - W]_{\mathrm{N, PN}}. \qquad (6.37)$$

We stress that the possibility of eliminating the PN contributions to R_1 and R_2 rested on the fact that *no interaction term does appear at this order in the expressions* $\partial Q/\partial v_{\tau j}$ and $\partial \rho/\partial v_{\tau j}$. This in turn is *a direct consequence of the assumptions made about the static nonrelativistic limit of the canonical potential and the validity of the "causality" conditions at the PN order.*

Finally, let us note that from the discussion of the preceding section and from Eqs. (6.36) it follows in particular that

$$\{x_{1i}, x_{1j}\} = \{x_{2i}, x_{2j}\} = 0,$$
$$\{x_{1i}, x_{2j}\} = O(1/c^4),$$
$$\{x_{\tau i}, p_{\tau' j}'\} = \delta_{\tau\tau'}\delta_{ij} + O(1/c^4),$$
$$\{p_{\tau i}', p_{\tau' j}'\} = O(1/c^4).$$

(6.38)

These equations show that the expressions p_1', p_2' be-

have as true momenta canonically conjugated to the physical variables $\mathbf{x}_1, \mathbf{x}_2$ up to the PN approximation.

Using now Eqs. (6.36), the *approximately relativistic Lagrangian* (6.37) assumes the explicit form

$$L = -mc^2 + L^{(0)}(\mathbf{x}_1, \mathbf{x}_2, \mathbf{v}_1, \mathbf{v}_2)$$
$$+ \frac{1}{c^2}L^{(1)}(\mathbf{x}_1, \mathbf{x}_2, \mathbf{v}_1, \mathbf{v}_2), \qquad (6.39)$$

where

$$L^{(0)} = \tfrac{1}{2}m_1\mathbf{v}_1^2 + \tfrac{1}{2}m_2\mathbf{v}_2^2 - U^{(0)}(r) \qquad (6.40)$$

and

$$L^{(1)} = \tfrac{1}{8}m_1\mathbf{v}_1^4 + \tfrac{1}{8}m_2\mathbf{v}_2^4 + \frac{1}{2}\left[\left(\frac{1}{m^2}(m_1\mathbf{v}_1 + m_2\mathbf{v}_2)^2\right.\right.$$
$$+ \frac{2\mu}{m}(\mathbf{v}_1 - \mathbf{v}_2)^2\right)U^{(0)}(r) + \left(\frac{1}{m^2}(\mathbf{r}(m_1\mathbf{v}_1 + m_2\mathbf{v}_2))^2\right.$$
$$- 2(\mathbf{r}\cdot\mathbf{v}_1)(\mathbf{r}\cdot\mathbf{v}_2)\right)\frac{1}{r}\frac{dU^{(0)}}{dr} - \left(\frac{1}{m}(\mathbf{v}_1 - \mathbf{v}_2)\cdot(m_1\mathbf{v}_1 + m_2\mathbf{v}_2)\right)$$
$$\times \Omega^{(1)}(r) - \left(\frac{1}{m}\mathbf{r}\cdot(\mathbf{v}_1 - \mathbf{v}_2)\mathbf{r}\cdot(m_1\mathbf{v}_1 + m_2\mathbf{v}_2)\right)\frac{1}{r}\frac{d\Omega^{(1)}}{dr}$$
$$- \left((\mathbf{v}_1 - \mathbf{v}_2)^2\Phi(r) + \frac{2\mu}{m}(\mathbf{v}_1 - \mathbf{v}_2)^2\omega^{(1)}(r)\right)$$
$$- \left((\mathbf{r}\cdot(\mathbf{v}_1 - \mathbf{v}_2))^2\Psi(r) + \frac{2\mu}{m}(\mathbf{r}\cdot(\mathbf{v}_1 - \mathbf{v}_2))^2\frac{1}{r}\frac{d\omega^{(1)}}{dr}\right)$$
$$- \left(2\Xi(r) + \frac{2r^2}{m}(U^{(0)}(r) - \omega^{(1)}(r))\frac{1}{r}\frac{dU^{(0)}}{dr}\right)\bigg], \qquad (6.41)$$

where, in particular, the *gauge character* of $\lambda^{(1)}$ $= (1/m)\,\omega^{(1)}(\rho)\cdot\sigma$ and its connection with the renormalization of the canonical interaction potential is directly evident from the last three lines. Actually the *renormalized canonical potential* [see also Eq. (5.64) and Eq. (6.7)] at the order $1/c^2$ results

$$\tilde{U}^{(1)}(\mathbf{r}, \mathbf{v}_1 - \mathbf{v}_2) = U^{(1)}(\mathbf{r}, \mathbf{v}_1 - \mathbf{v}_2) + \{\lambda^{(1)}, (Mc^2)^{(0)}\}$$
$$= \tilde{\Xi}(r) + \tfrac{1}{2}(\mathbf{v}_1 - \mathbf{v}_2)^2\tilde{\Phi}(r) + \tfrac{1}{2}(\mathbf{r}\cdot(\mathbf{v}_1 - \mathbf{v}_2))^2\tilde{\Psi}(r)$$

(6.42)

with

$$\tilde{\Xi}(r) = \Xi(r) - \frac{r^2}{m}\omega^{(1)}(r)\frac{1}{r}\frac{dU^{(0)}}{dr},$$

$$\tilde{\Phi}(r) = \Phi(r) + \frac{2\mu}{m}\omega^{(1)}(r), \qquad (6.43)$$

$$\tilde{\Psi}(r) = \Psi(r) + \frac{2\mu}{m}\frac{1}{r}\frac{d\omega^{(1)}}{dr}.$$

It is a remarkable result to find that *all the approximately relativistic "Lagrangians" for two particles existing in the* literature *belong to the general form summarized by our Eqs.* (6.39)—(6.41). A very general class of approximately relativistic Lagrangians corresponding to a PN approximation of linear variational principles of the Fokker type for n particles has been characterized by Woodcock and Havas[22] (hereafter denoted by W—H), (see also Refs. 23). [When the present paper had already been submitted for publication, P. Havas has kindly brought to our knowledge the existence of further related work going on by him and co-workers. See in particular Ref. 35b.] The W—H "Lagrangian" for 2-particles contains *three additional*

functions of the instantaneous interparticle separation
$|\mathbf{r}| = |\mathbf{x}_1 - \mathbf{x}_2|$ *besides the static Newtonian potential*
[*instead of our four* $\Omega^{(1)}(r)$, $\widetilde{\Xi}(r)$, $\widetilde{\Phi}(r)$, $\widetilde{\Psi}(r)$] and it can
be written

$$\mathcal{L} = -mc^2 + \{\tfrac{1}{2}m_1\mathbf{v}_1^2 + \tfrac{1}{2}m_2\mathbf{v}_2^2 - g_1g_2V_{12}(r)\}$$
$$+ \frac{1}{c^2}\{\tfrac{1}{8}(m_1\mathbf{v}_1^4 + m_2\mathbf{v}_2^4) + I_{\mathrm{PN}}\}, \qquad (6.44)$$

where

$$I_{\mathrm{PN}} = \tfrac{1}{2}g_1g_2\Big((\mathbf{v}_1 \cdot \mathbf{v}_2)V_{12}(r) - (\mathbf{r} \cdot \mathbf{v}_1)(\mathbf{r} \cdot \mathbf{v}_2)\frac{1}{r}\frac{dV_{12}}{dr}$$
$$+ (\mathbf{v}_1 - \mathbf{v}_2)^2(V_{12}(r) + X_{12}(r)) + (\mathbf{r} \cdot (\mathbf{v}_1 - \mathbf{v}_2))^2 Y_{12}(r)$$
$$+ \mathbf{v}_1 \cdot (\mathbf{v}_1 - \mathbf{v}_2)W_{12}(r) + (\mathbf{r} \cdot \mathbf{v}_2)(\mathbf{r} \cdot (\mathbf{v}_1 - \mathbf{v}_2))$$
$$\times \frac{1}{r}\frac{dW_{12}}{dr}\Big). \qquad (6.45)$$

It is easily seen that the expressions (6.44), (6.45) can
be obtained from Eqs. (6.39)—(6.41) by means of the
following identifications [W—H emphasize the absence
of contributions of order $1/c$ in their "Lagrangian."
From our point of view it is clear that no odd power
of $1/c$ can appear in the theory merely because of
the form of the world line condition (3.4) and the as-
sumption (4.1).]:

$$g_1g_2V_{12}(r) = U^{(0)}_{[\kappa_1,\kappa_2]}(r),$$

$$g_1g_2W_{12}(r) = \frac{m_1 - m_2}{m}U^{(0)}_{[\kappa_1,\kappa_2]}(r) - \Omega^{(1)}_{[\kappa_1,\kappa_2]}(r),$$

$$g_1g_2X_{12}(r) = -\Phi_{[\kappa_1,\kappa_2]}(r) + \frac{m_1^2}{m^2}U^{(0)}_{[\kappa_1,\kappa_2]}(r) - \frac{m_2}{m}\Omega^{(1)}_{[\kappa_1,\kappa_2]}(r)$$
$$\qquad - \frac{2\mu}{m}\omega^{(1)}(r)$$
$$= -\widetilde{\Phi}_{[\kappa_1,\kappa_2]}(r) + \frac{m_1^2}{m^2}U^{(0)}_{[\kappa_1,\kappa_2]}(r) - \frac{m_2}{m}\Omega^{(1)}_{[\kappa_1,\kappa_2]}(r)$$

$$g_1g_2Y_{12}(r) = -\Psi_{[\kappa_1,\kappa_2]}(r) + \frac{m_1^2}{m^2}\frac{1}{r}\frac{dU^{(0)}_{[\kappa_1,\kappa_2]}(r)}{dr} \qquad (6.46)$$
$$\qquad - \frac{m_1}{m}\frac{1}{r}\frac{d\Omega^{(1)}_{[\kappa_1,\kappa_2]}}{dr}(r) - \frac{2\mu}{m}\frac{1}{r}\frac{d\omega^{(1)}(r)}{dr}$$
$$= -\widetilde{\Psi}_{[\kappa_1,\kappa_2]}(r) + \frac{m_1^2}{m^2}\frac{1}{r}\frac{dU^{(0)}_{[\kappa_1,\kappa_2]}}{dr}(r)$$
$$\qquad - \frac{m_1}{m}\frac{1}{r}\frac{d\Omega^{(1)}_{[\kappa_1,\kappa_2]}(r)}{dr},$$

while the form of $\Xi(r)$ is directly related to the choice
of the "gauge" function $\omega^{(1)}(r)$ because it must have the
effect of *eliminating the terms which are nonlinear in
the interaction* $U^{(0)}(r)$ from the general "Lagrangian"
(6.39)—(6.41) since they cannot appear in the W—H
"Lagrangian." The most simple possibilities are

$$\omega^{(1)}(r) \equiv U^{(0)}(r), \quad \Xi(r) \equiv 0, \qquad (6.47)$$

$$\omega^{(1)}(r) \equiv 0, \quad \Xi(r) = -\frac{r^2}{m}U^{(0)}(r)\frac{1}{r}\frac{dU^{(0)}}{dr}. \qquad (6.48)$$

Of course in gauge-independent terms we must have

$$\Xi(r) = -\frac{r^2}{m}U^{(0)}(r)\frac{1}{r}\frac{dU^{(0)}}{dr}. \qquad (6.49)$$

A further very interesting problem is to investigate
the conditions, if any, which the various interaction
potentials must satisfy at the PN approximation in order

that there exist one or more of the so called "*adjunct*"
fields, associated with the Lorentz invariant variational
principle, which mediate the interaction among the
particles. This problem has also been settled by Havas
et al. (see Refs. 22, 23, and 35) in the same context.
The conditions found in Ref. 22 amount to saying that
the "Lagrangian" (6.44), (6.45) must be in general an
expression of the form

$$\mathcal{L} = -mc^2 + \{\tfrac{1}{2}m_1\mathbf{v}_1^2 + \tfrac{1}{2}m_2\mathbf{v}_2^2\} + \frac{1}{c^2}\{\tfrac{1}{8}(m_1\mathbf{v}_1^4 + m_2\mathbf{v}_2^4)\}$$
$$- \sum_{l,k}g_1^{(l,k)}g_2^{(l,k)}V_{12}^{(l,k)}(r) + \frac{1}{c^2}\sum_{l,k}I_{\mathrm{PN}}^{(l,k)}, \qquad (6.50)$$

where each term $I_{\mathrm{PN}}^{(l,k)}$ has the same structure as the
PN term in (6.45) with coupling constants and potentials
depending on the nonnegative integer numbers l and k
and satisfying certain identities. The numbers l and k
specify the powers under which the relativistic in-
variants θ and η_τ ($\tau = 1, 2$), respectively, appear in the
generalized Fokker integrand and essentially charac-
terize the structure of the source of the "adjunct" field
associated to each interaction term $I_{\mathrm{PN}}^{(l,k)}$; in particular
$l + k$ gives the tensorial rank of this field. Woodcock and
Havas work under the further simplifying assumption
that the coupling constants and the functional form of the
$V_{12}^{(l,k)}(r)$'s are all the same, independently of the values
of l, k, precisely

$$V_{12}(r) = \sum_{l,k}V_{12}^{(l,k)}(r) = \sum_{l,k}a^{(l,k)}V_{12}(r), \qquad (6.51)$$

with $a^{(l,k)}$ constants such that $\sum_{l,k}a^{(l,k)} = 1$. The identi-
ties which the W—H "potentials" must satisfy are

$$W_{12}^{(l,k)}(r) \equiv 0,$$
$$X_{12}^{(l,k)}(r) = -(l+k)\,V_{12}^{(l,k)}(r), \qquad (6.52)$$
$$Y_{12}^{(l,k)}(r) = \frac{k(1-k)}{2k-1}\frac{1}{r}\frac{dV_{12}^{(l,k)}}{dr}.$$

Under these conditions the W—H "Lagrangian" takes
the form

$$\mathcal{L} = -mc^2 + \{\tfrac{1}{2}m_1\mathbf{v}_1^2 + \tfrac{1}{2}m_2\mathbf{v}_2^2 - g_1g_1V_{12}(r)\}$$
$$+ \frac{1}{c^2}\Big\{\tfrac{1}{8}(m_1\mathbf{v}_1^4 + m_2\mathbf{v}_2^4) + \tfrac{1}{2}g_1g_2\Big[\sum_{l,k}a^{(l,k)}$$
$$\times (1 - l - k)(\mathbf{v}_1 - \mathbf{v}_2)^2 + (\mathbf{v}_1 \cdot \mathbf{v}_2)\Big]V_{12}(r) - \Big[(\mathbf{r} \cdot \mathbf{v}_1)(\mathbf{r} \cdot \mathbf{v}_2)$$
$$+ \sum_{l,k}a^{(l,k)}\frac{k(k-1)}{2k-1}(\mathbf{r} \cdot (\mathbf{v}_1 - \mathbf{v}_2))^2\Big]\frac{1}{r}\frac{dV_{12}(r)}{dr}\Big\}. \qquad (6.53)$$

Finally we can read this result in terms of our interac-
tion potentials. Taking into account Eqs. (6.39)—(6.53)
we obtain the *gauge-independent conditions*

$$\Omega^{(1)}_{[\kappa_1,\kappa_2]}(r) = \frac{m_1 - m_2}{m}U^{(0)}_{[\kappa_1,\kappa_2]}(r)$$

$$\widetilde{\Xi}_{[\kappa_1,\kappa_2]}(r) = -\frac{r^2}{m}U^{(0)}_{[\kappa_1,\kappa_2]}(r)\frac{1}{r}\frac{dU^{(0)}_{[\kappa_1,\kappa_2]}(r)}{dr}$$

$$\widetilde{\Phi}_{[\kappa_1,\kappa_2]}(r) = \Big[\sum_{l,k}a^{(l,k)}(l+k) - 1 + \frac{3\mu}{m}\Big]U^{(0)}_{[\kappa_1,\kappa_2]}(r) \qquad (6.54)$$

$$\widetilde{\Psi}_{[\kappa_1,\kappa_2]}(r) = \Big[\sum_{l,k}a^{(l,k)}\frac{k(k-1)}{2k-1} + \frac{\mu}{m}\Big]\frac{1}{r}\frac{dU^{(0)}_{[\kappa_1,\kappa_2]}(r)}{dr},$$

where it will be noticed that all the "potentials" are
expressed only in terms of $U^{(0)}(r)$. [*Note*: the first

equation implies in particular that the "distinguishability potential" $\Omega^{(1)}_{[\kappa_1,\kappa_2]}(r)$ must be *symmetrical* in its dependence on all the individual physical parameters *which are not masses.*] This might be an indication that, in the case of the existence of *"adjunct" fields*, the whole interaction occurring in the Newtonian-like equations of motion could be expressed as a functional of the Newtonian static interaction.

The general "Lagrangian" (6.39)—(6.41) with the conditions (6.43), (6.54) contains all the known approximately relativistic Lagrangians which have been used in the literature for the case of *linear field theories*: in particular the well-known Darwin—Breit "Lagrangian"[16,18] for electrodynamics, the Bagge "Lagrangian"[32] and a "Lagrangian" for a scalar field[23]; we have listed in a table the corresponding expressions for the interaction "potentials" and the various parameters involved. It is interesting to note in these cases the role played by the "distinguishability potential" $\Omega^{(1)}_{[\kappa_1,\kappa_2]}(r)$ which, as we have seen, comes from the "causality" conditions at the PN order. Woodcock and Havas discuss also a further explicit example of a purely spacelike interaction which in their language is *asymmetric* $[W_{12}(r) \neq 0]$ and *does not correspond to the existence of associate "adjunct" fields*. Explicitly, in terms of their potentials, it is defined by

$$W_{12}(r) = V_{12}(r), \quad X_{12}(r) = -lV_{12}(r),$$

$$Y_{12}(r) = \frac{1}{r}\frac{dV_{12}(r)}{dr}. \tag{6.55}$$

It should be noted that this example, corresponding to being

$$\Omega^{(1)}_{[\kappa_1,\kappa_2]}(r) = -\frac{2m_2}{m} U^{(0)}_{[\kappa_1,\kappa_2]}(r) \tag{6.56}$$

for *our* "distinguishability potential," is not consistent with the symmetry condition we have assumed in Sec. 3 [see Eqs. (3.17) and (5.27)] *if the masses only are allowed as characteristic individual parameters for the particles.* It is interesting to observe, however, that this example fits well with our scheme if it is viewed as a particular numerical determination of a more general theory in which the particles are characterized by, e.g., two *"charges"* η_1, η_2 as additional attributes. Assuming [compare with the previous *Note* following Eq. (6.54).]

$$\Omega^{(1)}_{[m_1,\eta_1;m_2,\eta_2]}(r) = \frac{2m_1\eta_1 - 2m_2\eta_2}{m} U^{(0)}_{[m_1,\eta_1;m_2,\eta_2]}(r), \tag{6.57}$$

we have a theory which satisfies the symmetry condition (b) if $U^{(0)}_{[m_1,\eta_1;m_2,\eta_2]}(r)$ is taken to be symmetric under the exchange $m_1 \rightleftarrows m_2$, $\eta_1 \rightleftarrows \eta_2$. On the other hand the corresponding W—H "potentials" result in

$$g_1 g_2 W_{[m,\eta]_{12}}(r)$$

$$= \left(\frac{m_1 - m_2}{m} + \frac{2m_2\eta_2 - 2m_1\eta_1}{m}\right) U^{(0)}_{[m,\eta]}(r),$$

$$g_1 g_2 X_{m,\eta]_{12}}(r)$$

$$= -\tilde{\Phi}_{[m,\eta]}(r) + \frac{m_1^2 + 2m_2^2\eta_2 - 2m_1 m_2\eta_1}{m^2} U^{(0)}_{[m,\eta]}(r), \tag{6.58}$$

$$g_1 g_2 Y_{[m,\eta]_{12}}(r)$$

$$= -\tilde{\Psi}_{[m,\eta]}(r) + \frac{m_1^2(1 - 2\eta_1) + 2m_1 m_2\eta_2}{m^2}\frac{1}{r}\frac{dU^{(0)}_{[m,\eta]}(r)}{dr}.$$

While Eqs. (6.55), (6.56) are obtained through the numerical determination $\eta_1 = 0$, $\eta_2 = 1$, the theory is symmetric in the sense of W—H only for the choice $\eta_1 = \eta_2 = \frac{1}{2}$: these remarks should further clarify the meaning of condition (b).

As we have seen, conditions of the form (6.47), (6.48), i.e., (6.49), which are necessary to reduce our "Lagrangian" to the W—H "Lagrangian," rule out all theories which contain nonlinear terms in the Newtonian interaction. This means that the general "Lagrangian" (6.39)—(6.41) does contain theories of this kind. Actually there are two important examples of nonlinear theories in the PN approximation which are well-known in the literature and belong to our scheme, namely the Einstein—Infeld—Hoffman (EIH) "Lagrangian" which is deduced from general relativity in the slow motion approximation[17,26] and the "Lagrangian" derived in the same approximation by Bazański[14,24] for the general relativistic equations of motion of n charged particles. [Recall that all the gravitation experiments concerning the solar system can be accounted for by using a weak-field limit of any metric theory of gravitation adjusting suitable parameters. This corresponds to what is called the *parametrized post-Newtonian formalism* (PPN).] For two particles they are

$$L_{\text{EIH}} = -mc^2 + \left\{\frac{1}{2}m_1\mathbf{v}_1^2 + \frac{1}{2}m_2\mathbf{v}_2^2 + G\frac{m_1 m_2}{r}\right\}$$

$$+ \frac{1}{c^2}\left\{\frac{1}{8}(m_1\mathbf{v}_1^4 + m_2\mathbf{v}_2^4) + \frac{1}{2}\left[(3\mathbf{v}_1^2 + 3\mathbf{v}_2^2 - 7(\mathbf{v}_1 \cdot \mathbf{v}_2)\right.\right.$$

$$\left.\left. - \frac{(\mathbf{r} \cdot \mathbf{v}_1)(\mathbf{r} \cdot \mathbf{v}_2)}{r^2}\right) G\frac{m_1 m_2}{r} - G^2\frac{m_1 m_2 m}{r^2}\right]\right\}, \tag{6.59}$$

and

$$L_B = -mc^2 + \left\{\frac{1}{2}m_1\mathbf{v}_1^2 + \frac{1}{2}m_2\mathbf{v}_2^2 + G\frac{m_1 m_2}{r} - \frac{e_1 e_2}{r}\right\}$$

$$+ \frac{1}{c^2}\left\{\frac{1}{8}(m_1\mathbf{v}_1^4 + m_2\mathbf{v}_2^4) + \left[\frac{1}{2}\left((\mathbf{v}_1 \cdot \mathbf{v}_2) + \frac{(\mathbf{r} \cdot \mathbf{v}_1)(\mathbf{r} \cdot \mathbf{v}_2)}{r^2}\right)\frac{e_1 e_2}{r}\right.\right.$$

$$+ \frac{1}{2}\left(3\mathbf{v}_1^2 + 3\mathbf{v}_2^2 - 7(\mathbf{v}_1 \cdot \mathbf{v}_2) - \frac{(\mathbf{r} \cdot \mathbf{v}_1)(\mathbf{r} \cdot \mathbf{v}_2)}{r^2}\right)G\frac{m_1 m_2}{r}$$

$$\left.\left. - \frac{1}{2}G^2\frac{m_1 m_2 m}{r^2} - G\left(\frac{me_1 e_2}{r^2} + \frac{m_1 e_2^2 + m_2 e_1^2}{2r^2}\right)\right]\right\}$$

$$= L_{\text{EIH}} + \frac{e_1 e_2}{r} + \frac{1}{c^2}\left\{\frac{1}{2}\left[(\mathbf{v}_1 \cdot \mathbf{v}_2) + \frac{(\mathbf{r} \cdot \mathbf{v}_1)(\mathbf{r} \cdot \mathbf{v}_2)}{r^2}\right]\frac{e_1 e_2}{r}\right.$$

$$\left. - G\left(\frac{me_1 e_2}{r^2} + \frac{m_1 e_2^2 + m_2 e_1^2}{2r^2}\right)\right\},$$

respectively.

In the Table we have listed the values of the parameters and the identifications of the various interaction potentials which relate also these two "Lagrangians" to the general form (6.39)—(6.41). The Bazański "Lagrangians" provides an example of a nonlinear superposition of field-related interactions with different tensor ranks and different coupling constants. In particular the gravitational part of both the EIH and

Bażański interactions corresponds to an *"adjunct" field* which is a mixture of a scalar and a tensor field of rank 2 (see the discussion contained in Refs. 22, 23). It is seen from the Table that for the Bażański case it is necessary to enlarge the conditions (6.54) [which correspond to the W—H case with the assumption (6.51)] by allowing for superposition of terms corresponding to the Newtonian interactions with different coupling constants in $\tilde{\Phi}(r)$ and $\tilde{\Psi}(r)$.

In conclusion we stress the great generality of the action-at-a-distance approach in the Hamiltonian form we have discussed in this paper. In connection with the existence of *approximately relativistic Lagrangians* for the Hamiltonian theory we should again emphasize the basic role played by the assumption (6.6) of a static nonrelativistic limit for the canonical interaction, which allows the construction of *PN physical canonical variables*. At the same time *if the Newtonian interaction* $U^{(0)}(r)$ *is not identically zero, the "causality" arguments given in Sec. 5 prevent the existence of approximately relativistic Lagrangians reproducing also terms of the order* $1/c^4$ *of the equations of motion*, i.e., the terms containing the radiative corrections. This makes clear the intrinsic limitations of the *true* Lagrangian formalism within the framework of relativistic particle dynamics. On the other hand it is always possible in principle to calculate all the $1/c^2$ expansion terms of $U(\rho, \pi)$ and $\Lambda(\rho, \pi)$ if a theory is given in the Newtonian-like form. This should be possible for instance in the case of electrodynamics by means of a Lagrange expansion in the sense of Kerner[26] for the Lorentz force term in the relativistic equations of motion. This procedure should provide in turn a canonical scheme for the theory.

7. CONCLUDING REMARKS

We conclude the paper with a number of remarks on the results obtained and some additional considerations.

Up to now we have discussed the "causality" conditions having in mind their formal quantum analog. We want now to show that the nonvanishing of the Poisson brackets $\{x_{1i}, x_{2j}\}$ can be related to some peculiar features of the Newtonian-like equations of motion.

From the formal point of view, the solution of Eqs. (6.4) corresponding to prescribed values of x_1, x_2, v_1, v_2 given at $t = 0$ does exist and is unique, just as it happens in nonrelativistic dynamics. Therefore we can say in this literal sense that the relativistic system composed by two particles interacting at a distance has just six degrees of freedom as its Newtonian counterpart and the theory possesses the so called *finitely predictive* character. [It is true that the number of degrees of freedom of the exact relativistic theory is unknown in general.[33] We agree, however, with Kerner's attitude which is not to speculate about the whole set of possible mathematical solutions but to consider as *physical motions* those which are analytically contiguous to the free motions (see also Ref. 34). Note that in the present work we have also implicitly assumed that the solutions became solutions of the nonrelativistic equations as $c \to \infty$.[33]] It must be stressed, however, as observed

by Havas in a very lucid discussion,[35a] that a relevant difference exists between the relativistic and the non-relativistic case. In the case of Eqs. (6.4) assuming certain values of the physical variables x_1, x_2, v_1, v_2 to have been *observed at time* $t = 0$, one can surely predict the future behavior of the system (and also infer its past behavior). It is not possible, however, to modify the dynamical variables of either particle arbitrarily without unavoidably perturbing also the dynamical variables of the other one at the same considered time if the space separation of the particles is such that the interaction is appreciable. Actually, let us consider for instance a situation in which the particles have been prepared in a certain way in the far past and then they have been allowed to evolve freely under their mutual action. At the instant $t = 0$ they will assume positions and velocities x_1', x_2', v_1', v_2'. Then, let us try to modify position and/or velocity of, say, the particle 1 leaving the far past conditions unchanged. In the nonrelativistic case this could be simply achieved by exerting a strong force on particle 1 during a short time interval Δt just before $t = 0$; for instance by allowing a third particle with very short-range interaction to pass near this particle during Δt. It is easily seen that this mechanism cannot work in the relativistic case since it is not possible to modify F_1 (or f_1) without modifying simultaneously also F_2 (or f_2): in other words a *situation in which a third particle interacts with particle 1 without interacting also with particle 2 cannot be realized, independently of the range of the forces involved*. In fact, once the equations of motion have been written in the form (6.27), it is clear that a modification of a certain portion of the world line 1 necessarily induces a modification of the world line 2 along the whole stretch which is spacelike with respect to the points of the modified part of world line 1. Therefore also F_2 (or f_2) in Eqs. (6.4) [or (6.5)] must be changed.

The above situation can be better understood from the canonical standpoint if we generalize our description to a three-particle system. The Hamiltonian for this system can be written for instance in the form

$$H = c\sqrt{M^2c^2 + \mathbf{P}^2}, \tag{7.1}$$

with

$$\begin{aligned} Mc^2 &= c\sqrt{m_1^2c^2 + \pi_1^2} + c\sqrt{m_2^2c^2 + \pi_2^2} \\ &\quad + c\sqrt{m_3^2c^2 + \pi_3^2} + U(\rho_{12}, \pi_{12}, \rho_3, \pi_3) \\ &= c\sqrt{M_{12}^2c^2 + \pi_3^2} + c\sqrt{m_3^2c^2 + \pi_3^2} + U(\rho_{12}, \pi_{12}, \rho_3, \pi_3), \end{aligned} \tag{7.2}$$

and

$$M_{12}c^2 = c\sqrt{m_1^2c^2 + \pi_{12}^2} + c\sqrt{m_2^2c^2 + \pi_{12}^2}. \tag{7.3}$$

Here the variable $\mathbf{P}$ denotes the total momentum of the system $\mathbf{P} = \mathbf{p}_1 + \mathbf{p}_2 + \mathbf{p}_3$; π_1, π_2, π_3 denote the momenta of particles 1, 2, 3 relative to the center-of-mass of the cluster $(1-2)$; π_3 reduces to the value of $\mathbf{p}_3$ in the center-of-momentum frame $(\mathbf{P} \equiv o)$ for $U \equiv 0$; π_{12} denotes the relative momentum of particles 1 and 2 and reduces to the value of $\mathbf{p}_1$ or $-\mathbf{p}_2$ in the center-of-mass system of the cluster $(1-2)$ $(\mathbf{p}_1 + \mathbf{p}_2 \equiv o)$ for $U \equiv 0$; ρ_3 and ρ_{12} are the variables canonically conjugated to π_3 and π_{12}, respectively; (the remaining canonical variable $\mathbf{Q}$, con-

jugated to **P**, is the usual canonical center-of-mass defined in terms of the Poincaré generators). It is now possible to choose U in such a way that particles 1 and 2 interact while particle 3 is free. As a matter of fact if we set

$$U \equiv U_{12} = c\sqrt{[M_{12} + (1/c^2)h_{12}(\rho_{12}, \pi_{12})]^2 c^2 + \pi_3^2}$$
$$- c\sqrt{M_{12}^2 c^2 + \pi_3^2}, \qquad (7.4)$$

we obtain

$$Mc^2 = c\sqrt{[M_{12} + (1/c^2)h_{12}]^2 c^2 + \pi_3^2} + c\sqrt{m_3^2 c^2 + \pi_3^2}. \qquad (7.5)$$

Then the expression of the canonical generators turns out to be formally identical to the free case for *two* particles apart from the replacement of π with π_3, the first particle mass with the effective mass of the cluster $(1-2)$ $M_{12} + (1/c^2)h_{12}$, and the introduction of a spin $\mathbf{s}_{12} = \rho_{12} \wedge \pi_{12}$ for particle 1. Consequently the generators can be split in a sum of two terms, the first of which depending only on the variables of the cluster and the second only on the variables of particle 3. Should we construct in this situation the physical variables $\mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_3$ and the corresponding Newtonian-like equations following the procedure used for two particles, we would find $\mathbf{f}_3 \equiv o$ while $\mathbf{f}_1$ and $\mathbf{f}_2$ would depend on the variables $\mathbf{r}_{12} \equiv \mathbf{x}_1 - \mathbf{x}_2, \mathbf{v}_1, \mathbf{v}_2$ only. Corresponding to any other choice of the form of U, the coordinates of the three particles enter the canonical generators in such a complicated structure that the forces $\mathbf{f}_1, \mathbf{f}_2, \mathbf{f}_3$ must depend on the variables $\mathbf{r}_{12}, \mathbf{r}_{13}, \mathbf{r}_{23}$ and the velocities of all of the particles. Therefore if we have interacting pairs $(1-2)$ and $(1-3)$, particles 2 and 3 must also interact. Thus, for instance, if we choose [according to Eq. (7.4)]

$$U = U_{12} + U_{13}, \qquad (7.6)$$

U_{13} affects the expression of $\mathbf{f}_2$ besides that of $\mathbf{f}_1$.

In order to make clear the connection between the above situation and the lack of "causality" of the theory in the sense used so far in this paper, let us assume that the perturbation of the world line 1 can be described as the result of an infinitesimal canonical transformation generated by a function $G(\mathbf{x}_1, \mathbf{x}_2, \mathbf{v}_1, \mathbf{v}_2)$. This would be a realistic mechanism for instance if the mass of the third particle in Eqs. (7.2) is very large compared with m_1 and m_2 and the perturbation is much smaller than U_{12} or effective only for a short time interval. If we now want that the perturbation modifies, say, $\mathbf{x}_1$ and $\mathbf{v}_1$ without affecting $\mathbf{x}_2$ and $\mathbf{v}_2$, the function G must satisfy the relations

$$\{G, \mathbf{x}_1\} \neq 0, \quad \{G, \mathbf{v}_1\} \neq 0, \qquad (7.7)$$

$$\{G, \mathbf{x}_2\} = \{G, \mathbf{v}_2\} = 0, \qquad (7.8)$$

which imply in turn, via the Jacobi identity,

$$\{G, \{x_{2i}, v_{2j}\}\} = 0. \qquad (7.9)$$

On the other hand from the "localizability" conditions and the Jacobi identity it follows that

$$\{x_{2i}, v_{2j}\} = \{x_{2j}, v_{2i}\}, \qquad (7.10)$$

with the consequence that the independent expressions of the form $\{x_{2i}, v_{2j}\}$ are just six in number. Were now these quantities functions of $\mathbf{x}_2$ and $\mathbf{v}_2$ only, as it happens

to be for free particles and in the nonrelativistic case (precisely functions of only $\mathbf{v}_2$ and constants, respectively), the system (7.8) would be Jacobian and would admit six independent solutions. Consequently it would be possible to modify the components of $\mathbf{x}_1$ and $\mathbf{v}_1$ in a completely independent way. In the actual relativistic case with interaction, however, as it is apparent from the expressions we have explicitly calculated, the quantities $\{x_{2i}, v_{2j}\}$ do depend on $\mathbf{x}_1, \mathbf{v}_1$ besides $\mathbf{x}_2, \mathbf{v}_2$ and even more, for given values of $\mathbf{x}_2, \mathbf{v}_2$, they are independent functions of the former variables. Then from Eqs. (7.8) it follows that

$$\{G, \mathbf{x}_1\} = \{G, \mathbf{v}_1\} = 0 \qquad (7.11)$$

(i. e., also $G \equiv 0$), which contradicts Eqs. (7.7). That this result rests specifically on the lack of "causality" is apparent from the fact that the vanishing of the expressions $\{x_{1i}, x_{2j}\}$ combined with the covariance of the vectors $\mathbf{x}_1, \mathbf{x}_2$ entails the simultaneous vanishing of the mixed expressions $\{x_{1i}, v_{2j}\}, \{x_{2i}, v_{1j}\}, \{v_{1i}, v_{2j}\}$. These last identities in turn, combined with Jacobi identity, would imply that $\{x_{2i}, v_{2j}\}$ are functions of only $\mathbf{x}_2$ and $\mathbf{v}_2$ in force of the Poisson bracket relations

$$\{x_{1k}, \{x_{2i}, x_{2j}\}\} = -\{x_{2i}, \{v_{2j}, x_{1k}\}\} - \{v_{2j}, \{x_{1k}, x_{2i}\}\},$$
$$\{v_{1k}, \{x_{2i}, v_{2j}\}\} = -\{x_{2i}, \{v_{2j}, v_{1k}\}\} - \{v_{2j}, \{v_{1k}, x_{2i}\}\}. \qquad (7.12)$$

These considerations shed some light into the inner physical mechanism of the zero-interaction theorem. If we assume the point of view of looking at the Hamiltonian formulation as obtained from a Newtonian-like differential theory in the form studied by Hill and Kerner, as outlined in the Appendix, we should not forget the circuitous route along which this latter can be deduced from a manifestly covariant action at a distance theory of the general Van Dam—Wigner type,[27] under the assumption of analytical connection of the actual world lines to the free particle motions.[26] It is apparent that, while the physical limitations put on the initial value problem by the *hereditary character* of relativistic dynamics are obscured in the transition from the complete integro-differential formulation to the instantaneous differential form, they reappear in a meaningful way within the Hamiltonian formulation as a lack of "dynamical independence" or "causality" of the particles positions, an effect which manifests itself necessarily at the order $1/c^4$, i. e., at the order of the radiation effects.[36] This "dynamical dependence" no longer occurs for the canonical variables Q, ρ since they are in some sense only *mean* variables in the region where the interaction is important. On the other hand, as we have seen, it is always possible to save the "localizability" of the particles positions, a fact which in the spirit of the above discussion, should correspond to the existence of an intrinsic meaning of the world lines themselves. Since we have explicitly given a method for the construction of physical position variables (in terms of the basic canonical variables) which satisfy the world line condition to any order in $1/c^2$, our results should apparently disprove the widespread opinion according to which the zero-interaction theorem prevents the existence of invariant world lines within a relativistic Hamiltonian framework.[10,5,6,7] At the same time while we agree that the physical implications of

1490   J. Math. Phys., Vol. 17, No. 9, August 1976

M. Pauri and G.M. Prosperi   1490

any theory in which the world line condition would be violated are rather obscure,[35,41] we can claim that abandoning the identification of the canonical variables as physical positions of the particles does not render the relativity principle necessarily vacuous.[11,13,14,8]

A second point to be examined is the structure which the canonical generators of the Poincaré group assume in terms of the physical variables x_1, x_2, v_1, v_2. From Eqs. (2.4), (2.5), (2.7), (4.4)–(4.6), taking into account Eqs. (6.14) we have

$$\mathbf{P} = \mathbf{P}^{(0)} + (1/c^2)\mathbf{P}^{(1)} + \cdots$$

$$= \{m_1\mathbf{v}_1 + m_2\mathbf{v}_2\} + \frac{1}{c^2}\left\{\frac{1}{m}\left[\tfrac{1}{2}m_1 v_1^2 + \tfrac{1}{2}m_2 v_2^2 + U^{(0)}(r)\right]\right.$$

$$\times (m_1\mathbf{v}_1 + m_2\mathbf{v}_2) + \left[\frac{m_1 - m_2}{m}\mathbf{r}\cdot(\mathbf{v}_1 - \mathbf{v}_2)\right.$$

$$- \frac{1}{m}\mathbf{r}\cdot(m_1\mathbf{v}_1 + m_2\mathbf{v}_2)\bigg]\frac{1}{r}\frac{dU^{(0)}}{dr}\mathbf{r}$$

$$+ \left[\frac{\mu}{m}(\mathbf{v}_1 - \mathbf{v}_2)\cdot(m_1\mathbf{v}_1 + m_2\mathbf{v}_2) - \frac{\mu}{2m}(m_1 - m_2)\right.$$

$$\times (\mathbf{v}_1 - \mathbf{v}_2)^2\bigg](\mathbf{v}_1 - \mathbf{v}_2) - \tfrac{1}{2}\Omega^{(1)}(r)(\mathbf{v}_1 - \mathbf{v}_2)$$

$$\left. - \frac{1}{2}\frac{1}{r}\frac{d\Omega^{(1)}}{dr}\mathbf{r}\cdot(\mathbf{v}_1 - \mathbf{v}_2)\mathbf{r}\right\} + \cdots, \tag{7.13}$$

$$H = mc^2 + H^{(0)} + (1/c^2)H^{(1)} + \cdots$$

$$= mc^2 + \{\tfrac{1}{2}m_1 v_1^2 + \tfrac{1}{2}m_2 v_2^2 + U^{(0)}(r)\}$$

$$+ \frac{1}{c^2}\left\{\frac{3}{8}(m_1 v_1^4 + m_2 v_2^4) + \frac{1}{2}\left[\frac{1}{m^2}(m_1\mathbf{v}_1 + m_2\mathbf{v}_2)^2\right.\right.$$

$$+ \frac{2\mu}{m}(\mathbf{v}_1 - \mathbf{v}_2)^2\bigg]U^{(0)}(r) + \frac{1}{2}\bigg[\frac{1}{m^2}(\mathbf{r}\cdot(m_1\mathbf{v}_1 + m_2\mathbf{v}_2))^2$$

$$- 2(\mathbf{r}\cdot\mathbf{v}_1)(\mathbf{r}\cdot\mathbf{v}_2)\bigg]\frac{1}{r}\frac{dU^{(0)}}{dr} - \frac{1}{2m}(\mathbf{v}_1 - \mathbf{v}_2)\cdot(m_1\mathbf{v}_1$$

$$+ m_2\mathbf{v}_2)\cdot\Omega^{(1)}(r) - \frac{1}{2m}\mathbf{r}\cdot(\mathbf{v}_1 - \mathbf{v}_2)\mathbf{r}\cdot(m_1\mathbf{v}_1 + m_2\mathbf{v}_2)$$

$$\times \frac{1}{r}\frac{d\Omega^{(1)}}{dr} - \tfrac{1}{2}(\mathbf{v}_1 - \mathbf{v}_2)^2\widetilde{\Phi}(r) - \tfrac{1}{2}(\mathbf{r}\cdot(\mathbf{v}_1 - \mathbf{v}_2))^2\widetilde{\Psi}(r)$$

$$\left. + \frac{1}{m}r^2 U^{(0)}(r)\frac{1}{r}\frac{dU^{(0)}}{dr} + \widetilde{\Xi}(r)\right\} + \cdots, \tag{7.14}$$

$$\mathbf{J} = \mathbf{J}^{(0)} + (1/c^2)\mathbf{J}^{(1)} + \cdots$$

$$= \{m_1\mathbf{x}_1 \wedge \mathbf{v}_1 + m_2\mathbf{x}_2 \wedge \mathbf{v}_2\} + \frac{1}{c^2}\left\{\tfrac{1}{2}m_1 v_1^2 \mathbf{x}_1 \wedge \mathbf{v}_1\right.$$

$$+ \tfrac{1}{2}m_2 v_2^2 \mathbf{x}_2 \wedge \mathbf{v}_2 + \frac{1}{m}[m_1\mathbf{x}_1 \wedge \mathbf{v}_1 + m_2\mathbf{x}_2 \wedge \mathbf{v}_2$$

$$+ \mu\mathbf{r} \wedge (\mathbf{v}_1 - \mathbf{v}_2)]U^{(0)}(r) - \frac{1}{2m}[\mathbf{r} \wedge (m_1\mathbf{v}_1 + m_2\mathbf{v}_2)$$

$$+ (m_1\mathbf{x}_1 + m_2\mathbf{x}_2) \wedge (\mathbf{v}_1 - \mathbf{v}_2)]\Omega^{(1)}(r)$$

$$\left. - \widetilde{\Phi}(r)\mathbf{r} \wedge (\mathbf{v}_1 - \mathbf{v}_2)\right\} + \cdots, \tag{7.15}$$

$$\mathbf{K} = \mathbf{K}^{(0)} + (1/c^2)\mathbf{K}^{(1)} + \cdots$$

$$= -\{m_1\mathbf{x}_1 + m_2\mathbf{x}_2\} - \frac{1}{c^2}\left\{\tfrac{1}{2}m_1 v_1^2 \mathbf{x}_1 + \tfrac{1}{2}m_2 v_2^2 \mathbf{x}_2\right.$$

$$\left. + \frac{1}{m}(m_1\mathbf{x}_1 + m_2\mathbf{x}_2)U^{(0)}(r) - \tfrac{1}{2}\Omega^{(1)}(r)\mathbf{r}\right\} + \cdots. \tag{7.16}$$

These expressions also represent the *ten approximate constants of motion which follow from the approximately relativistic Lagrangian* (6.39)–(6.41) via the standard route of the Noether theorem (see Refs. 23 and 25). Of course, from our point of view, they are but the first approximation to the ten *exact* constants of the motion which we knew from the beginning. We can see in particular that the linear and the angular momentum do not have their simple standard kinematical form in terms of x_1, x_2, v_1, v_2, a fact which is but another manifestation of the zero-interaction theorem[37] and is related to the physical circumstance that within the instantaneous formulation there must appear terms representing linear and angular momentum which are being transferred between the particles at the given instant. In this connection, moreover, it is interesting to see the form assumed by the center of mass defined in Ref. 1. We have

$$\mathbf{R} = \left\{\frac{m_1\mathbf{x}_1 + m_2\mathbf{x}_2}{m}\right\} + \frac{1}{c^2}\left\{\frac{1}{2m}[\mu(v_1^2 - v_2^2) - \Omega^{(1)}(r)]\mathbf{r}\right\} + \cdots$$

$$= \left[\frac{m_1}{m} + \frac{1}{c^2}\frac{1}{2m}\left(\mu(v_1^2 - v_2^2) - \Omega^{(1)}_{[\kappa_1,\kappa_2]}(r)\right)\right]\mathbf{x}_1$$

$$+ \left[\frac{m_2}{m} + \frac{1}{c^2}\frac{1}{2m}\left(\mu(v_2^2 - v_1^2) - \Omega^{(1)}_{[\kappa_2,\kappa_1]}(r)\right)\right]\mathbf{x}_2 + \cdots, \tag{7.17}$$

where the role of the "distinguishability" potential $\Omega^{(1)}_{[\kappa_1,\kappa_2]}(r)$ should be noticed and the fact that $\mathbf{R}$, at least to the PN approximation, does retain a Newtonian-like structure,[23]

$$\mathbf{R}_{N,PN} = \sum_{\tau=1}^{2} (M_\tau/M)\,\mathbf{x}_\tau, \tag{7.18}$$

with generalized masses given by

$$\frac{M_\tau}{M} = \frac{m_\tau}{m} + \frac{1}{c^2}\frac{1}{2m}\left(\mu(v_\tau^2 - v_{\tau'}^2) - \Omega^{(1)}_{[\kappa_\tau,\kappa_{\tau'}]}(r)\right), \quad (\tau \neq \tau'). \tag{7.19}$$

Note also the different roles that the "potentials" play in the expressions given above. In particular, the nonlinear term $\widetilde{\Xi}(r)$ occurs only in the Hamiltonian. Finally, it is apparent that in the asymptotic limit $r \to \infty$ the standard free particle expressions of the generators in terms of physical variables are recovered only in the case of short-range forces even if, as we have seen, *the accelerations themselves vanish in the same limit.* In the case of long-range forces, asymptotic interaction terms survive in general in the expressions of the generators of the *homogeneous Lorentz subgroup* (relativistic angular momentum). For instance, under the conditions (6.54) for the existence of associated "*adjunct*" fields, we find

$$\mathbf{J}_{r \to \infty} \Rightarrow \mathbf{J}^{(free)}_{asympt} + \mathbf{J}^{(int)}_{asympt},$$
$$\mathbf{K}_{r \to \infty} \Rightarrow \mathbf{K}^{(free)}_{asympt} + \mathbf{K}^{(int)}_{asympt}, \tag{7.20}$$

with

$$\mathbf{J}^{(0)(int)}_{asympt} = 0,$$

$$\mathbf{J}^{(1)(int)}_{asympt} = U^{(0)}(r)\left\{\mathbf{r} \wedge \left[\left(1 - \frac{2\mu}{m} - \sum_{l,k} a^{(l,k)}(l+k)\right)(\mathbf{v}_1 - \mathbf{v}_2)\right.\right.$$

$$-\frac{m_1-m_2}{2m^2}(m_1\mathbf{v}_1+m_2\mathbf{v}_2)\Big]-(m_1\mathbf{x}_1+m_2\mathbf{x}_2)$$

$$\wedge\frac{m_1-m_2}{2m^2}(\mathbf{v}_1-\mathbf{v}_2)+\frac{m_1}{m}\mathbf{x}_1\wedge\mathbf{v}_1+\frac{m_2}{m}\mathbf{x}_2\wedge\mathbf{v}_2\Big\},$$

$$\mathbf{K}^{(0)(int)}_{asympt}=0,$$

$$\mathbf{K}^{(1)(int)}_{asympt}=-\tfrac{1}{2}U^{(0)}(r)(\mathbf{x}_1+\mathbf{x}_2),\qquad(7.21)$$

which do not vanish asymptotically if $U^{(0)}(r)\sim 1/r$. These results are in agreement with the conclusions drawn by Van Dam and Wigner and Cooke[38] in the general case of manifestly covariant theories based on action principles of the Fokker type. In particular we recover the fact that the asymptotic interaction terms are a $1/c^2$ effect. It is interesting to find that for long-range forces the "causality" conditions themselves are not satisfied asymptotically at the order $1/c^4$, a fact which must be interpreted on the basis of our discussion on the initial value problem. For example if Eqs. (6.54) are satisfied we, find

$$\{x_{1i},x_{2j}\}^{(2)}$$

$$=-\frac{1}{m^2\mu}\left\{(\mathbf{r}\cdot(m_1\mathbf{v}_1+m_2\mathbf{v}))\left[\frac{3}{2}+\sum_{l,k}a^{(l,k)}(l+k)\right]\right\}$$

$$\times U^{(0)}(r)\delta_{ij}+\frac{1}{m^3}\Big\{(\mathbf{r}\cdot(m_1\mathbf{v}_1+m_2\mathbf{v}_2))\frac{m}{\mu}$$

$$\times\left[\frac{1}{2}-\sum_{l,k}a^{(l,k)}\frac{k(k-1)}{2k-1}\right]-\mathbf{r}\cdot(\mathbf{v}_1-\mathbf{v}_2)(m_1-m_2)$$

$$\times\left[\frac{m}{\mu}-\frac{3}{2}+\frac{m}{\mu}\sum_{l,k}a^{(l,k)}\frac{k(k-1)}{2k-1}\right]\Big\}\frac{1}{r}\frac{dU^{(0)}}{dr}r_ir_j$$

$$+\frac{1}{m^2}\Big\{\frac{m_1-m_2}{2\mu}\Big[\tfrac{1}{2}-\sum_{l,k}a^{(l,k)}(l+k)\Big]U^{(0)}(r)\Big\}$$

$$\times(r_i(v_{1j}-v_{2j})+r_j(v_{1i}-v_{2i}))+\frac{1}{m^2}$$

$$\times\Big\{\frac{1}{2}\left[\frac{(m_1-m_2)^2}{2m_1m_2}+\frac{6\mu}{m}+\left(2-\frac{m}{\mu}\right)\right.$$

$$\times\sum_{l,k}a^{(l,k)}(l+k)\Big]U^{(0)}(r)\Big\}(r_i(v_{1j}-v_{2j})$$

$$-r_j(v_{1i}-v_{2i}))+\frac{1}{\mu^2}\frac{\partial^2\Lambda^{(2)}}{\partial(\mathbf{v}_1-\mathbf{v}_2)_i\,\partial(\mathbf{v}_1-\mathbf{v}_2)_j},\qquad(7.22)$$

which fails to vanish asymptotically for generic values of the masses no matter how $\Lambda^{(2)}(\rho,\pi)$ is chosen (for an exact one-dimensional case see R.N. Hill).[13] Clearly, all of these asymptotic features are strictly related to the fact that for long-range forces the physical position vectors $\mathbf{x}_1,\mathbf{x}_2$ do not approach asymptotically in a strict sense the free particle expressions $\mathbf{q}_1,\mathbf{q}_2$, as already mentioned at the end of Sec. 5.

A third point to be examined is the consideration of possible superluminal velocities of the particles. It has been observed by Currie[13] that the relativistic invariance of the differential equations of motion does not prevent by itself alone the possibility that the particles' velocity happens to exceed the velocity of light. It is worth noticing that this possibility can be ruled out in our context under simple regularity assumptions on the interaction potentials $U(\rho,\pi)$ and $\Lambda(\rho,\pi)$. Actually, should the velocity of one particle, say $\mathbf{v}_1$, exceed the

velocity of light, there would exist a reference frame in which $\mathbf{v}_1$ becomes infinite. On the other hand, since $\mathbf{v}_1=\{\mathbf{x}_1,H\}$, this can happen only for values of $\rho,\pi,\mathbf{P}$ for which $\mathbf{x}_1$ or H are singular. Now the singularities of the $\mathbf{x}_r$'s and H are related to the singularities of U and Λ. If, for instance we assume that these functions are singular only for $\rho=0$, $\mathbf{v}_1$ may happen to exceed c only if a reference frame exists in which ρ vanishes at a certain time. Since, however, the phase-space trajectory of ρ as viewed in different reference frames differs only for a change in the time scale and for a space rotation [see Ref. 1, Eqs. (4.30)—(4.36)], the vanishing of ρ has an intrinsic meaning and thus must occur in every reference frame if it does in a particular one. Consequently the velocity of the particles may possibly exceed c for a set of initial conditions having zero measure at most. Let us note that the assumption on the singularities of U, Λ are just met, at least in the PN approximation, in all the specific examples discussed in Sec. 6.

A final point is to consider the limiting situation in which the mass of one of the particles, say m_2, goes to infinity. It is easy to find that $\mathbf{x}_1$ and $\mathbf{x}_2$ become causal in this limit, independently of the form of the functions U,Λ (and λ), and we recover the nonrelativistic expressions

$$\mathbf{x}_1\Rightarrow\mathbf{Q}+\rho,\quad\mathbf{x}_2\Rightarrow\mathbf{Q}.\qquad(7.23)$$

In the same limit, putting $\mathbf{P}=0$, $\mathbf{Q}=0$, the Hamiltonian becomes

$$H=[m_2c^2]+m_1c^2+\left\{\frac{\mathbf{p}_1^2}{2m_1}+U^{(0)}(r)\right\}+\frac{1}{c^2}\left\{-\frac{1}{8m_1^3}\mathbf{p}_1^4\right.$$

$$\left.+\frac{1}{2m_1^2}\mathbf{p}_1^2\tilde{\Phi}(r)+\frac{1}{2m_1^2}(\mathbf{r}\cdot\mathbf{p}_1)^2\tilde{\Psi}(r)\right\}+\cdots.\qquad(7.24)$$

It is interesting to distinguish here the various possibilities which occur in the case of an associated "adjunct" field corresponding to different tensorial ranks. For example if $k=l=0$ (scalar field) we have

$$\tilde{\Phi}(r)\equiv-U^{(0)}(r),\quad\tilde{\Psi}(r)\equiv0,\qquad(7.25)$$

i.e.,

$$H=[m_2c^2]+m_1c^2+\left\{\frac{1}{2m_1}\mathbf{p}_1^2+U^{(0)}(r)\right\}$$

$$-\frac{1}{c^2}\left\{\frac{1}{8m_1^3}\mathbf{p}_1^4+\frac{1}{2m_1^2}\mathbf{p}_1^2U^{(0)}(r)\right\}+\cdots,\qquad(7.26)$$

which is but the PN expansion of

$$H=[m_2c^2]+c\sqrt{[m_1+(1/c^2)U^{(0)}(r)]^2c^2+\mathbf{p}_1^2},\qquad(7.27)$$

typical of an external scalar field. On the other hand if $k=0$, $l=1$ (vector field) we have

$$\tilde{\Phi}(r)\equiv0,\quad\tilde{\Psi}(r)\equiv0,\qquad(7.28)$$

i.e.,

$$H=[m_2c^2]+m_1c^2+\left\{\frac{1}{2m_1}\mathbf{p}_1^2+U^{(0)}(r)\right\}$$

$$+\frac{1}{c^2}\left\{-\frac{1}{8m_1^3}\mathbf{p}_1^4\right\}+\cdots,\qquad(7.29)$$

which is the PN expansion of

$$H=[m_2c^2]+c\sqrt{m_1^2c^2+\mathbf{p}_1^2}+U^{(0)}(r),\qquad(7.30)$$

typical of a static external vector field.

1492   J. Math. Phys., Vol. 17, No. 8, August 1976

M. Pauri and G.M. Prosperi   1492

TABLE I. Approximately Relativistic Lagrangians

$$\mathcal{L} = -mc^2 + \{\tfrac{1}{2} m_1 \mathbf{v}_1{}^2 + \tfrac{1}{2} m_2 \mathbf{v}_2{}^2 - U^{(0)}(r)\}$$
$$+ (1/c^2)\{\tfrac{1}{8} m_2 \mathbf{v}_2{}^4 + I_{PN}\}$$

Post-Newtonian Interaction Terms

A. Linear case

1. Havas and Stachel "Lagrangian" for scalar field[23]

$$I_{PN} = \tfrac{1}{2}\left[(\mathbf{v}_1{}^2 + \mathbf{v}_2{}^2 - \mathbf{v}_1 \cdot \mathbf{v}_2)\, U^{(0)}(r) - (\mathbf{r} \cdot \mathbf{v}_1)(\mathbf{r} \cdot \mathbf{v}_2)\frac{1}{r}\frac{dU^{(0)}}{dr}\right]$$

$k = 0$, $l = 0$: single "adjunct" scalar field

$$\Omega^{(1)}(r) = (m_1 - m_2)/m\, U^{(1)}(r),$$

$$\widetilde{\Xi}(r) \equiv -\frac{r^2}{m} U^{(0)}(r)\frac{1}{r}\frac{dU^{(0)}}{dr},$$

$$\widetilde{\Phi}(r) \equiv (3\mu/m - 1)\, U^{(0)}(r),$$

$$\widetilde{\Psi}(r) \equiv \frac{\mu}{m}\frac{1}{r}\frac{dU^{(0)}}{dr},$$

2. Bagge[32]

$$I_{PN} = \tfrac{1}{2}\left[(\mathbf{v}_1 \cdot \mathbf{v}_2)\, U^{(0)}(r) - (\mathbf{r} \cdot \mathbf{v}_1)(\mathbf{r} \cdot \mathbf{v}_2)\frac{1}{r}\frac{dU^{(0)}}{dr}\right]$$

$k = 0$, $l = 0$: single "adjunct" vector field

$$\Omega^{(1)}(r) \equiv \left(\frac{m_1 - m_2}{m}\right)U^{(0)}(r)$$

$$\widetilde{\Xi}(r) \equiv -\frac{r^2}{m} U^{(0)}(r)\frac{1}{r}\frac{dU^{(0)}}{dr},$$

$$\widetilde{\Phi}(r) \equiv (3\mu/m)\, U^{(0)}(r),$$

$$\widetilde{\Psi}(r) \equiv \frac{\mu}{m}\frac{1}{r}\frac{dU^{(0)}}{dr}.$$

A particular case of 2 is the electromagnetic "Lagrangian", 2a.

2a. Darwin—Breit[16,18]

$$U^{(0)}(r) = e_1 e_2/r,$$

B. Nonlinear case

3. Einstein—Infeld—Hoffmann[17]

$$I_{PN} = -\tfrac{1}{2}G^2 m_1 m_2 m/r^2 + \tfrac{1}{2}\left[3\mathbf{v}_1{}^2 + 3\mathbf{v}_2{}^2 - 7\mathbf{v}_1 \cdot \mathbf{v}_2 - \frac{(\mathbf{r} \cdot \mathbf{v}_1)(\mathbf{r} \cdot \mathbf{v}_2)}{r^2}\right]G\frac{m_1 m_2}{r},$$

$k = 0$, $l = 0, 2$, $a^{(0,0)} = -1$, $a^{(2,0)} = 2$: superposition of "adjunct" scalar and tensor field of rank 2,

$$U^{(0)}(r) = -Gm_1 m_2/r,$$

$$\Omega^{(1)}(r) \equiv [(m_1 - m_2)/m]U^{(0)}(r) = G(m_1 - m_2)\mu/r,$$

$$\widetilde{\Xi}(r) \equiv -\frac{r^2}{m} U^{(0)}(r)\frac{1}{r}\frac{dU^{(0)}}{dr} = G^2(m + 2\mu)m_1 m_2/2r^2$$

$$= (1/m + 1/2\mu)[U^{(0)}(r)]^2,$$

$$\widetilde{\Phi}(r) \equiv 3(1 + \mu/m)U^{(0)}(r),$$

$$\widetilde{\Psi}(r) \equiv \frac{\mu}{m}\frac{1}{r}\frac{dU^{(0)}}{dr}.$$

4. Bażański[24]

$$I_{PN} = -\tfrac{1}{2}G^2\frac{m_1 m_2 m}{r^2} - G\left(\frac{me_1 e_2}{r^2} + \frac{m_1 e_2{}^2 + m_2 e_1{}^2}{2r^2}\right)$$

$$+ \tfrac{1}{2}[3\mathbf{v}_1{}^2 + 3\mathbf{v}_2{}^2 - 7\mathbf{v}_1 \cdot \mathbf{v}_2 - (\mathbf{r} \cdot \mathbf{v}_1)(\mathbf{r} \cdot \mathbf{v}_2)/r^2]Gm_1 m_2/r$$

$$+ \tfrac{1}{2}[(\mathbf{v}_1 \cdot \mathbf{v}_2) + (\mathbf{r} \cdot \mathbf{v}_1)(\mathbf{r} \cdot \mathbf{v}_2)/r^2]e_1 e_2/r,$$

TABLE I. (Continued)

$k = 0$, $l = 1 : a_E{}^{(1,0)} = 1$,

$k = 0$, $l = 0, 2 : a_G{}^{(0,0)} = -1$, $a_G{}^{(2,0)} = 2$:

superposition of "adjunct" scalar, vector and tensor field of rank 2

$$U^{(0)}(r) = U_G^{(0)}(r) + U_E^{(0)}(r) = Gm_1 m_2/r + e_1 e_2/r.$$

$$\Omega^{(1)}(r) = \frac{m_1 m_2}{m} U^{(0)}(r) = -G(m_1 - m_2)\mu/r + e_1 e_2(m_1 - m_2)/mr,$$

$$\widetilde{\Xi}(r) = \left(\frac{1}{m} + \frac{1}{2\mu}\right)[U_G^{(0)}(r)]^2 + \frac{1}{m}\{U_E^{(0)}(r)\}^2$$

$$+ 2\left(\frac{1}{m} - \frac{1}{2\mu} + \frac{m_2 e_1{}^2 - m_1 e_2{}^2}{2me_1 e_2}\right)U_G^{(0)}(r)U_E^{(0)}(r),$$

$$\widetilde{\Phi}(r) = \frac{3\mu}{m}U_E^{(0)}(r) + 3\left(1 + \frac{\mu}{m}\right)U_G^{(0)}(r),$$

$$\widetilde{\Psi}(r) = \frac{\mu}{m}\frac{1}{r}\frac{d}{dr}[U_G^{(0)}(r) + U_E^{(0)}(r)].$$

To conclude, we want only to stress that the results obtained in the present paper tend to emphasize the main significance of the fully canonical description summarized by Eqs. (2.4), (2.5) and (2.7) as that of an *asymptotical description* which gives only *mean information* on the detailed space—time development of the system and from which, for example, the "distinguishability potential" $\Omega^{(1)}(r)$ has disappeared. From this point of view, if we are given a theory in the Newtonian-like form, we are naturally lead to an *asymptotic canonical quantization* along the standard route or possibly on a more geometrical basis[39] in the spirit of an S-matrix formulation of the quantum theory. A less ambitious program could be for instance to perform a Bohr—Sommerfeld quantization[40] in a given approximation for some of the theories contained in our general "Lagrangian" (6.39)—(6.41), such as the gravitational "Lagrangians" of EIH or Bażański.

APPENDIX

We assume we have a direct interaction theory in the form discussed by Kerner and Hill (see Refs. 19 and, in general, 14 and 20). Then the equations of motion for the two-particle system are expressed as an analytic first order system of differential equations in terms of the physical positions x_1 and x_2 and velocities v_1 and v_2 which, from the present point of view, play the role of fundamental variables. We pretend now that the state of the system can be completely specified by a set of canonical coordinates Q, P, ρ, π. By means of the method of Hamiltonization proposed in Ref. 19, it is possible to prescribe Poisson bracket relations among these variables in such a way that the transformations of the Poincaré group are canonical transformations. Then we can construct the infinitesimal generators as functions of the variables x_1, x_2, v_1, v_2 and, consequently, the variables of the scheme B (see Ref. 1) for the given realization. Taking into account Eqs. (2.11), we obtain

$$Q = Q(x_1, x_2, v_1, v_2), \quad P = P(x_1, x_2, v_1, v_2). \tag{A1}$$

1493   J. Math. Phys., Vol. 17, No. 8, August 1976

M. Pauri and G.M. Prosperi   1493

Then we set by definition

$$P_4(x_1, x_2, v_1, v_2) \equiv S_z(x_1, x_2, v_1, v_2) = \pi_\varphi,$$

$$Q_4(x_1, x_2, v_1, v_2) \equiv \arctan \frac{S_y(x_1, x_2, v_1, v_2)}{S_x(x_1, x_2, v_1, v_2)}$$

$$\equiv \arctan \frac{\cos\varphi\, \pi_\theta - \sin\varphi\, \cot\theta\, \pi_\varphi}{-\sin\varphi\, \pi_\theta - \cos\varphi\, \cot\theta\, \pi_\varphi}$$

$$P_5(x_1, x_2, v_1, v_2) \equiv S(x_1, x_2, v_1, v_2) \tag{A2}$$

$$= \pi_\theta^2 + \frac{1}{\sin^2\theta}\, \pi_\varphi^2$$

$$Q_5(x_1, x_2, v_1, v_2)$$

$$+ f[S(x_1, \ldots, v_2), Q_6(x_1, \ldots, x_2), P_6(x_1, \ldots, v_2)]$$

$$= \arctan \frac{\pi_\theta \tan\theta}{S},$$

where f is an arbitrary function of its arguments. By comparing such equations with the scheme B of the rotation group for the single particle realization (see Ref. 3), we see that the expressions obtained by inversion of Eq. (A2)

$$\varphi = \varphi(x_1, x_2, v_1, v_2), \quad \pi_\varphi = \pi_\varphi(x_1, x_2, v_1, v_2),$$

$$\theta = \theta(x_1, x_2, v_1, v_2), \quad \pi_\theta = \pi_\theta(x_1, x_2, v_1, v_2), \tag{A3}$$

are canonical variables and it follows that

$$S_x = -\sin\varphi\, \pi_\theta - \cos\varphi\, \cot\theta\, \pi_\varphi,$$

$$S_y = \cos\varphi\, \pi_\theta - \sin\varphi\, \cot\theta\, \pi_\varphi, \tag{A4}$$

$$S_z = \pi_\varphi.$$

Finally, let us construct two quantities, $A[S(x_1, \ldots, v_2), Q_6(x_1, \ldots, v_2), P_6(x_1, \ldots, v_2)]$ and $B[S(x_1, \ldots, v_2), Q_6(x_1, \ldots, v_2), P_6(x_1, \ldots, v_2)]$, which have zero Poisson bracket with

$$Q_5(x_1, \ldots, v_2)$$

$$+ f[S(x_1, \ldots, v_2), Q_6(x_1, \ldots, v_2), P_6(x_1, \ldots, v_2)]. \tag{A5}$$

As a consequence, A and B have zero Poisson bracket also with $\varphi, \theta, \pi_\varphi, \pi_\theta$. Therefore we can set $\rho = g(A, B)$ (g arbitrary) and construct finally a canonically conjugate variable $\pi_\rho = \Psi(A, B)$. In conclusion we have a set of internal canonical variables such that

$$\rho = \rho(x_1, x_2, v_1, v_2), \quad \pi = \pi(x_1, x_2, v_1, v_2), \tag{A6}$$

$$S = \rho \wedge \pi. \tag{A7}$$

It is clear that in the above derivation there are two elements of arbitrariness, namely the choice of the functions f and g. Now, in order that the state of the system can be specified by the values of the variables Q, P, ρ, π, it is necessary that the system of equations (A1) and (A6) is *globally invertible*, i.e., establishes a one-to-one correspondence between the two 12-dimensional Euclidean spaces coordinatized by the Cartesian variables Q, P, ρ, π and x_1, x_2, v_1, v_2, respectively. *This must be compatible with the asymptotic conditions.* Precisely, in order that ρ and $x_1 - x_2$ can be identified in the center-of-mass system for large separation of the particles, it is necessary that Eqs. (2.2) and (2.3) are satisfied asymptotically when q_1 and q_2 are replaced by x_1 and x_2 and p_1, p_2 by their usual relativistic free

particle expressions in terms of v_1 and v_2. In conclusion, starting from the Newtonian differential theory, our basic assumptions would be satisfied if the above conditions could be met for some appropriate choice of the functions f and g. That this is not a trivial point is shown by the following considerations. One could directly identify the variables $Q_4(x_1, x_2, v_1, v_2), \ldots,$ $P_6(x_1, x_2, v_1, v_2)$ with the corresponding rhs expressions in the scheme B for the free particles.[1] Then, by inverting these relations, one could construct the variables $\rho(x_1, x_2, v_1, v_2)$ and $\pi(x_1, x_2, v_1, v_2)$ in such a way that also the function Mc has the free particle expression

$$Mc = \pi_{10} + \pi_{20}. \tag{A8}$$

However, were Eqs. (A1) and (A6) globally invertible in this case, with the asymptotic identification of the particle positions with q_1 and q_2 simultaneously satisfied, we would have a quite peculiar kind of interaction without scattering effects and possibility of bound states of motion. (We think of bound states as of trajectories which are bounded in the variable $|x_{1_{c.m.}} - x_{2_{c.m.}}|$ and thus in $\rho_{c.m.}$ from $t = -\infty$ to $t = +\infty$. Alternatively, bounded motions can be defined in terms of an action-angle variables formulation for the internal dynamics.) This situations is strictly analogous to that of the standard nonrelativistic Hamilton—Jacobi theory where any system of interacting particles can be reduced to the free particle form by means of suitable canonical transformations; these transformations, however, are defined only locally in the phase space.

[1] M. Pauri and G.M. Prosperi, J. Math. Phys. 15, 1503 (1975).

[2] M. Pauri and G.M. Prosperi, J. Math. Phys. 7, 366 (1966).

[3] M. Pauri and G.M. Prosperi, J. Math. Phys. 8, 2256 (1967).

[4] M. Pauri and G.M. Prosperi, J. Math. Phys. 9, 1146 (1968).

[5] B. Bakamjian and L.H. Thomas, Phys. Rev. 92, 1300 (1953). See also P.M.A. Dirac, Rev. Mod. Phys. 21, 392 (1949); L.H. Thomas, Rev. Mod. Phys. 17, 182 (1945); Phys. Rev. 85, 868 (1952); B. Bakamjian, Phys. Rev. 121, 1849 (1961).

[6] L.L. Foldy, Phys. Rev. 122, 275 (1961); T.F. Jordan, A.J. Mac Farlane and E.C.G. Sudarshan, Phys. Rev. B 133, 487 (1964); T.F. Jordan, J. Math. Phys. 5, 1345 (1964); F. Coester, Helv. Phys. Acta 38, 7 (1965); N.D. Son and J. Sucher, Phys. Rev. 153, 1496 (1967); D. Avison, Phys. Rev. 154, 1583 (1967); D.G. Currie and T.F. Jordan, *Lectures in Theoretical,* edited by W.E. Brittin and A.O. Barut (Gordon and Breach, New York, 1968), Vol. X-A, p. 91.

[7] J.E. Krizan and P. Havas, Phys. Rev. 128, 2916 (1962); P. Havas, *Proceedings of the International Symposium on Statistical Mechanics and Thermodynamics,* edited by J. Meixner (North-Holland, Amsterdam, 1965).

[8] R. Fong and J. Sucher, J. Math. Phys. 5, 456 (1964).

[9] D.G. Currie, J. Math. Phys. 4, 1470 (1963); D.G. Currie, T.F. Jordan, and E.C.G. Sudarshan, Rev. Mod. Phys. 35, 350 (1963); J.T. Cannon and T.F. Jordan, J. Math. Phys. 5, 299 (1964); H. Leutwyler, Nuovo Cimento 37, 560 (1965); T.F. Jordan, Phys. Rev. 166, 1308 (1968).

[10] H. van Dam and E.P. Wigner, Phys. Rev. B 138, 1576 (1965).

[11] A. Peres, Phys. Rev. Lett. 27, 1666 (1971); 28, 392 (1972); P. Droz-Vincent, Nuovo Cimento B 12, 1 (1972); see the implications of Ref. 8 for quantum mechanics.

[12] E.H. Kerner, J. Math. Phys. 6, 1218 (1965); R.N. Hill,

J. Math. Phys. **8**, 1756 (1967); C.S. Shukre and T.F. Jordan Jordan, J. Math. Phys. **13**, 868 (1972).

[13]D.G. Currie, Phys. Rev. **142**, 817 (1966); Bull. Am. Phys. Soc. **11**, 356 (1966); Ann. Phys. (N.Y.) **36**, 104 (1966); E.H. Kerner, Phys. Rev. Lett. **16**, 667 (1966); Bull. Am. Phys. Soc. **11**, 373 (1966); D.G. Currie and T.F. Jordan, Phys. Rev. Lett. **16**, 1210 (1966); J.G. Wray "Two Particle Forces for Relativistic Newtonian Equations of Motion," Preprint NYO-3829-37, July 1969; R.N. Hill, J. Math. Phys. **11**, 1918 (1970); see also for alternative viewpoints R. Arens, J. Math. Phys. **7**, 1341 (1966); Arch. Ration. Mech. Anal. **47**, 255 (1972); "Motions of Relativistic Hamiltonian Interactions," Preprint, University of California, May 1973; "Models of Relativistic Hamiltonian Interactions," Preprint, University of California, November 1973; "Models of Hamiltonian several particle interactions," Preprint, University of California, July 1974; R. Arens and D.G. Babbitt, Pac. J. Math. **28**, 243 (1969); L. Bel, Ann. Inst. Henri Poincaré **14A**, 189 (1971); **18A**, 57 (1973); H.P. Kunzle, "Poincaré invariant Hamiltonian two-particle systems and the relativistic Kepler problem," Preprint, University of Alberta, Canada, 1973; J. Math. Phys. **15**, 1033 (1974).

[14]P. Droz-Vincent, Lett. Nuovo Cimento **1**, 839 (1969); Phys. Scr. **2**, 129 (1970).

[15]P. Havas and J. Plebański, Bull. Am. Phys. Soc. **5**, 433 (1961).

[16]C.G. Darwin, Philos. Magn. **39**, 537 (1920); J.L Anderson, *Principles of Relativity Physics* (Academic, New York, 1967), Chap. 7, Sec. 7—11.

[17]A. Einstein, L. Infeld, and B. Hoffmann, Ann. Math. **39**, 66 (1938); V. Fock, *The Theory of Space Time and Gravitation* (Pergamon, London, 1959), Chap. VI, Sec. 78; L. Infeld and J. Plebański, *Motion and Relativity* (Pergamon, London, 1960), Chap. III, Sec. 3, Chap. IV, Sec. 2.

[18]G. Breit, Phys. Rev. **34**, 553 (1929); **51**, 248 (1937); **53**, 153 (1938); H. Primakoff and T. Holstein, Phys. Rev. **55**, 1218 (1939); R.B. Hoffman and P. Havas, Phys. Rev. B **140**, 1162 (1965); J.E. Krizan, Phys. Rev. A **140**, 1155 (1965); **152**, 136 (1966); **177**, 376 (1969); E. Breitenberger, Am. J. Phys. **36**, 505 (1968); H. Osborne, Phys. Rev. **176**, 1514 (1968); **176**, 1523 (1968); F.E. Close and H. Osborne, Phys. Rev. D **2**, 2127 (1970); T.E. Dengler and J.E. Krizan, Phys. Rev. A **2**, 2388 (1970); R.A. Krajcik and L.L. Foldy, Phys. Rev. Lett. **24**, 545 (1970).

[19]R.N. Hill, J. Math. Phys. **8**, 1756 (1967); R.N. Hill and E. Kerner, Phys. Rev. Lett. **17**, 1156 (1966).

[20]R.N. Hill, J. Math. Phys. **8**, 201 (1967); E. Kerner, J. Math. Phys. **9**, 222 (1968).

[21]L. Landau and E. Lifchitz, "Théorie du Champ," (Editions de la Paix, Moscou, 1967), Chap. 8, Sec. 65, Chap. 11, Sec. 105.

[22]H.W. Woodcock and P. Havas, Phys. Rev. D **6**, 3422 (1972).

[23]P. Havas and J. Stachel, Phys. Rev. **185**, 1636 (1969); F.J. Kennedy, Am. J. Phys. **40**, 63 (1972); P. Havas, in *Problems in the Foundations of Physics*, edited by M. Bunge (Springer, New York, 1971), p. 31.

[24]S. Bażański, Acta Phys. Pol. **15**, 423 (1957); **15**, 363 (1966); *Recent Developments in General Relativity* (Pergamon, London, 1962), p. 137.

[25]S. Chandrasekhar and G. Contopoulos, Proc. R. Soc. London A **298**, 123 (1967).

[26]E. Kerner, J. Math. Phys. **3**, 35 (1962); Phys. Rev. **125**, 2184 (1962).

[27]H. Tetrode, Z. Phys. **10**, 317 (1922); A.D. Fokker, Z. Phys. **58**, 386 (1929); Physica **9**, 33 (1929); **12**, 145 (1932); J.A. Wheeler and R.P. Feynman, Rev. Mod. Phys. **17**, 157 (1945); **21**, 425 (1949); R.M.F. Houtappel, H. van Dam, and E.P. Wigner, Rev. Mod. Phys. **37**, 595 (1965); A. Schild and J.A. Schlosser, J. Math. Phys. **6**, 1299 (1965); **9**, 913 (1968); F.Rohrlich, *Classical Charged Particles* (Addison-Wesley, Reading, Mass., 1965), Chap. 7; J.V. Narlikar, Proc. Cambridge Philos. Soc. **64**, 1071 (1968); A. Katz, J. Math. Phys. **10**, 1929 (1969); **10**, 2215 (1969); J.E. Krizan, Phys. Rev. D **1**, 2772 (1970); E.C.G. Sudarshan, CPT-55-AEC, September 1970, Preprint, University of Texas at Austin; W.C. Schieve, A. Rosenblum, and P. Havas, Phys. Rev. D **6**, 1501 (1972); D **6**, 1522 (1972); P. Cordero and G.C. G.C. Ghirardi, J. Math. Phys. **14**, 815 (1973).

[28]R. Courant and D. Hilbert, *Methods of Mathematical Physics* (Interscience, New York, 1962), Vol. II, Chap. II, Sec. 2.

[29]See for example K. Kumar, J. Math. Phys. **6**, 1923 (1965); **6**, 1928 (1965); R.M. Wilcox, J. Math. Phys. **8**, 962 (1967).

[30]S. Lardera, E. Montaldi, and G.M. Prosperi, Nuovo Cimento A **70**, 103 (1970).

[31]S. Gasiorowicz and M.A. Ruderman, Phys. Rev. **110**, 261 (1958); H. Ekstein, Phys. Rev. **117**, 1590 (1960).

[32]E. Bagge, Z. Naturforsch. **1**, 361 (1946).

[33]R.D. Driver, Ann. Phys. (N.Y.) **21**, 122 (1963); R.D. Driver and M.J. Norris, Ann. Phys. (N.Y.) **42**, 347 (1967); R.D. Driver, Phys. Rev. **178**, 2051 (1969); C.M. Andersen and H.C. von Baeyer, Ann. Phys. (N.Y.) **60**, 67 (1970); Phys. Rev. D **5**, 802 (1972); D **5**, 2470 (1972).

[34]H.J. Bhabha, Phys. Rev. **70**, 759 (1946).

[35](a) P. Havas, in *Proceedings of the 1964 International Congress of Logic, Methodology and Philosophy of Science* (North-Holland, Amsterdam, 1965), p. 347; Sinthèse **18**, 75 (1968); "Causality and Relativistic Dynamics" presented at the *Symposium on Causality and Physical Theories*, Wayne State University, May 1973 (AIP Conference Proceedings No. 16, edited by W.B. Rolnick). (b) J. Stachel and P. Havas, "Invariance of Approximately Relativistic Hamiltonians," Preprint, 1975.

[36]E. Breitenberger, Am. J. Phys. **36**, 505 (1968).

[37]A.N. Beard and R. Fong, Phys. Rev. **182**, 1397 (1967); R.A. Rudd and R.N. Hill, J. Math. Phys. **11**, 2704 (1970); T.F. Jordan and C.S. Shukre, "Momentum and angular momentum in relativistic classical particle mechanics," Preprint, NYO 3829-79 September 1971; D.G. Currie and C.S. Shukre, Phys. Rev. D **5**, 799 (1972); R. Arens, "Momentum of relativistic 4-particle interactions," Preprint, University of California, March 1972.

[38]H. van Dam and E.P. Wigner, Phys. Rev. **142**, 838 (1966); J.H. Cooke, Phys. Rev. **174**, 1602 (1968).

[39]E. Onofri and M. Pauri, J. Math. Phys. **13**, 533 (1972); Lett. Nuovo Cimento **3**, 35 (1972).

[40]A. Degasperis, Phys. Rev. D **3**, 273 (1971).

Asymptotic radiation from spinning charged particles

J. Cohn and H. Wiebe*†

Department of Physics and Astronomy, University of Oklahoma, Norman, Okla. 73069
(Received 27 May 1975; final revised manuscript received 5 November 1975)

A manifestly covariant expression for the asymptotic energy–momentum and angular momentum emitted by a charged spinning particle in arbitrary motion is found. A center of energy theorem is discussed and it is also shown that, for spinning particles, the radiation rate is not an invariant.

I. INTRODUCTION

The purpose of the present work is, in part, to correct a previous paper written by one of us,[1] and to extend the considerations of that work. The present work is the first of a two part series dealing, first, with primarily the calculation of the asymptotic linear and angular momentum emitted by an arbitrarily moving spinning classical (nonquantized) charged particle and, in the second part, with a derivation of the consequent equations of motion of such a particle.

The following discussion will deal, successively, with the construction of the electromagnetic field tensor corresponding to an arbitrarily moving spinning charge—the evaluation of the corresponding stress—tensor—and consequently, with the construction of the expressions for the asymptotically radiated linear and angular momenta.

Finally, a center of energy theorem will be briefly discussed, and it will also be shown that the energy radiation rate from a spinning charge is not an invariant.

The discussion will be manifestly covariant throughout.

II. EM FIELD TENSOR

Here, we take as the starting point in calculating the field tensor, $F^{\mu\nu}$, the known expression for the 4-vector potential A^{μ}, which is the sum of a nonspin part, $A^{\mu}_{\text{n.s.}}$, and a spin contribution A^{μ}_{s}, given by[2]

$$A^{\mu}_{\text{n.s.}} = -\frac{eV^{\mu}}{R^{\alpha}V_{\alpha}} \tag{II.1}$$

and

$$A^{\mu}_{s} = \frac{e}{2mR^{\alpha}V_{\alpha}}\frac{d}{d\tau}\left(\frac{M^{\mu\nu}R_{\nu}}{R^{\beta}V_{\beta}}\right), \tag{II.2}$$

where we are using the metric $g_{\alpha\beta} = \text{diag}(1,1,1,-1)$, and world length $d\tau^2 = -g_{\alpha\beta}dX^{\alpha}dX^{\beta}$; $R^{\alpha} = X^{\alpha} - Z^{\alpha}$ is the light vector, where X^{α} denotes the field event, and Z^{α} the retarded particle event; $V^{\mu} = dX^{\mu}/d\tau$; $M^{\mu\nu}$ is the moment tensor characterizing the particle, where we assume that the rest frame electric dipole moment of the particle is zero (so that $M^{\alpha\beta}V_{\beta} = 0$); m is the particle mass, and we take the speed of light to be one.

Carrying out the above differentiation involved in A^{μ}_{s} and utilizing the nonelectric nature of $M^{\mu\nu}$, we obtain

$$A^{\mu} = \frac{e}{\rho}V^{\mu} + \frac{e}{2m\rho^2}\left(\overset{\circ}{M}^{\mu\nu}R_{\nu} + \frac{M^{\mu\nu}R_{\nu}R^{\alpha}a_{\alpha}}{\rho} + \frac{M^{\mu\nu}\dot{R}_{\nu}}{\rho}\right), \tag{II.3}$$

where a dot over a symbol signifies differentiation with respect to τ, $a^{\mu} = dV^{\mu}/d\tau$, and $\rho = -R^{\alpha}V_{\alpha}$ and physically denotes the 3-distance between the field event and retarded charge event—in the momentary rest frame of the retarded charge.

Now, we wish to calculate the field tensor, $F^{\mu\nu} = \partial^{\mu}A^{\nu} - \partial^{\nu}A^{\mu}$. To facilitate this, we use the general relation[3]

$$\partial^{\nu}C^{\mu} = \overset{\circ}{C}^{\mu}\partial^{\nu}\tau = -\overset{\circ}{C}^{\mu}R^{\nu}/\rho \tag{II.4}$$

for any kinematical quantity C^{μ}.

We thus obtain

$$\partial^{\nu}A^{\mu}_{\text{n.s.}} = -(e/\rho^2)(a^{\mu}R^{\nu} + V^{\mu}\lambda^{\nu}) \tag{II.5}$$

and

$$\partial^{\nu}A^{\mu}_{s} = -(e/m\rho^3)(\tfrac{3}{2}M^{\mu\alpha}U_{\alpha}\lambda^{\nu} + \tfrac{1}{2}\overset{\circ}{M}^{\mu\alpha}U_{\alpha}R^{\nu}$$
$$-\tfrac{1}{2}M^{\mu\nu} + \overset{\circ}{M}^{\mu\alpha}R_{\alpha}\lambda^{\nu} + \tfrac{1}{2}\overset{\circ\circ}{M}^{\mu\alpha}R_{\alpha}R^{\nu})$$
$$-(e/m\rho^2)(\tfrac{3}{2}a_{U}M^{\mu\alpha}U_{\alpha}\lambda^{\nu} + \tfrac{1}{2}a_{U}\overset{\circ}{M}^{\mu\alpha}U_{\alpha}R^{\nu}$$
$$+\tfrac{1}{2}\overset{\circ}{a}_{U}M^{\mu\alpha}U_{\alpha}R^{\nu} - \tfrac{1}{2}a^2M^{\mu\alpha}U_{\alpha}R^{\nu} - \tfrac{1}{2}M^{\mu\alpha}U_{\alpha}a^{\nu}$$
$$-a_{U}M^{\mu\nu} - \tfrac{1}{2}a_{U}M^{\mu\alpha}a_{\alpha}R^{\nu} - \tfrac{1}{2}\overset{\circ}{M}^{\mu\nu}), \tag{II.6}$$

where $\lambda^{\alpha} \equiv U^{\alpha} + a_{U}R^{\alpha}$, $a_{U} \equiv a_{\sigma}U^{\sigma}$, $\overset{\circ}{a}_{U} \equiv \overset{\circ}{a}_{\sigma}U^{\sigma}$, and U^{μ} is defined by the relation, $R^{\mu} = \rho(U^{\mu} + V^{\mu})$.

Adding these results and antisymmetrizing then yields

$$F^{\mu\nu} = F^{\mu\nu}_{(-1)} + F^{\mu\nu}_{(-2)} + F^{\mu\nu}_{(-3)} \tag{II.7}$$

with

$$F^{\mu\nu}_{(-1)} \equiv -(e/m\rho^2)[\tfrac{3}{2}a^2_{U}M^{[\nu\alpha}U_{\alpha}R^{\mu]} + \tfrac{1}{2}a_{U}\overset{\circ}{M}^{[\nu\alpha}U_{\alpha}R^{\mu]}$$
$$+\tfrac{1}{2}\overset{\circ}{a}_{U}M^{[\nu\alpha}U_{\alpha}R^{\mu]} - \tfrac{1}{2}a^2M^{[\nu\alpha}U_{\alpha}R^{\mu]}$$
$$-\tfrac{1}{2}a_{U}M^{[\nu\alpha}a_{\alpha}R^{\mu]} + (a_{U}/\rho)\overset{\circ}{M}^{[\nu\alpha}R_{\alpha}R^{\mu]}$$
$$+(1/2\rho)\overset{\circ\circ}{M}^{[\nu\alpha}R_{\alpha}R^{\mu]} + ma^{[\nu}R^{\mu]} + ma_{U}V^{[\nu}R^{\mu]}], \tag{II.7a}$$

$$F^{\mu\nu}_{(-2)} \equiv -(e/m\rho^2)[\tfrac{3}{2}(a_{U}/\rho)M^{[\nu\alpha}U_{\alpha}R^{\mu]} + (1/2\rho)\overset{\circ}{M}^{[\nu\alpha}U_{\alpha}R^{\mu]}$$
$$+\tfrac{3}{2}a_{U}M^{[\nu\alpha}U_{\alpha}U^{\mu]} - \tfrac{1}{2}M^{[\nu\alpha}U_{\alpha}a^{\mu]}$$
$$+a_{U}M^{[\nu\mu]} + (1/\rho)\overset{\circ}{M}^{[\nu\alpha}R_{\alpha}U^{\mu]} + \overset{\circ}{M}^{[\nu\mu]} + mV^{[\nu}U^{\mu]}], \tag{II.7b}$$

$$F^{\mu\nu}_{(-3)} \equiv -(e/m\rho^3)(\tfrac{3}{2}M^{[\nu\alpha}U_{\alpha}U^{\mu]} + M^{\mu\nu}) \tag{II.7c}$$

and where the lower index $(-1, -2, -3)$ indicates the

order of ρ entering the expression—and further,
$$M^{[\nu\alpha}U_\alpha R^{\mu]} \equiv M^{\nu\alpha}U_\alpha R^\mu - M^{\mu\alpha}U_\alpha R^\nu$$ illustrates the meaning of the bracket notation.

In passing, we note that the above expression for $F^{\mu\nu}$ is in keeping with a theorem of Goldberg and Kerr.[4] According to this theorem, $F^{\mu\nu}$ must have the ρ-dependence indicated above, and in addition $F_{(-1)}^{\mu\nu}R_\nu = 0$ and $F_{(-2)}^{\mu\nu}R_\nu = AR^\mu$, A being a scalar. A brief calculation shows that, indeed, these latter two conditions are satisfied.

III. VALIDITY OF MOMENTA EXPRESSIONS

Here we construct the general expressions for the asymptotically radiated linear and angular momenta and demonstrate that they satisfy certain necessary conditions of definition.

We take the expression for the asymptotic radiated momenta to be of the same form as for a nonspinning charge. So we have[5]

$$dP_{\rm rad}^\mu = \lim_{\rho \to \infty} \int_{(\Delta\sigma)} \Theta^{\mu\nu} d\sigma_\nu \tag{III.1}$$

and

$$dJ_{\rm rad}^{\mu\nu} = \lim_{\rho \to \infty} \int_{(\Delta\sigma)} J^{\alpha\mu\nu} d\sigma_\nu \tag{III.2}$$

for the asymptotically radiated linear and angular momenta, respectively, where the limit procedure and $(\Delta\sigma)$ together denote the customary asymptotic spacelike annular hypersurface. Also, $\Theta^{\mu\nu}$ denotes the symmetric stress tensor (to be considered in detail presently) and

$$J^{\alpha\mu\nu} \equiv -(\Theta^{\alpha\mu}X^\nu - \Theta^{\alpha\nu}X^\mu) \tag{III.2'}$$

is the angular momentum "density."

In order for these definitions to be justified, however, it must be shown that they are independent of the orientation of the spacelike surface $(\Delta\sigma)$. We now turn to those considerations while referring to Fig. 1. Now, since $\Theta^{\mu\nu}$ and $J^{\alpha\mu\nu}$ both have vanishing 4-divergence in the 4-volume bounded by $\Delta\sigma_1$, $\Delta\sigma_2$ and the light-cone surfaces ΔC_1 and ΔC_2, Gauss' theorem implies that the integrals of $\Theta^{\mu\nu}$ and $J^{\alpha\mu\nu}$ over $\Delta\sigma_1$ and $\Delta\sigma_2$ are equal, provided that the integrals over ΔC_1 and ΔC_2 vanish, and this equality implies that the above definitions are, indeed, independent of the orientation of $(\Delta\sigma)$.
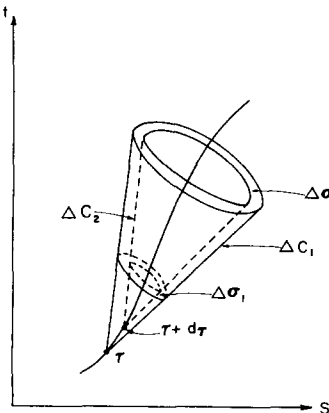
FIG. 1. Light-cone construction related to momentum emission.

To prove the point of concern then, we shall demonstrate that, in the limit as $\rho \to \infty$, the relevant integrals over ΔC_1 and ΔC_2 do vanish. We consider the linear and angular momentum cases separately:

For the linear case we shall prove the point by showing that

$$\lim_{\rho \to \infty} \int_{\Delta C} \Theta^{\mu\nu} R_\nu d^2\omega = 0, \tag{III.3}$$

where the integration is over either ΔC_1 or ΔC_2, and we note that the light-cone surface element is given by $R^\alpha d^2\omega$, where $d^2\omega$ is the invariant 2-element associated with the light cone, and is given by $\rho d\rho d\omega$, where $d\omega$ is an ordinary differential solid angle. Thus, $R_\nu d^2\omega$ depends quadratically on ρ.[6]

Now, $\Theta^{\mu\nu}$ is quadratic in the $F^{\alpha\beta}$, which in turn, depends on terms of order ρ^{-1}, ρ^{-2} and ρ^{-3} [see Eq. (II.7)]. Thus, we only need to consider that part of $\Theta^{\mu\nu}$ which depends on ρ^{-2}—since we are taking $\rho \to \infty$.

By a straightforward calculation we find that

$$\Theta_{(-2)}^{\mu\nu} = (1/4\pi)(F_{(-1)}^{\mu\alpha}F_{\alpha(-1)}^\nu + \tfrac{1}{4}g^{\mu\nu}F_{(-1)}^{\alpha\beta}F_{\alpha\beta(-1)}) \tag{III.4a}$$

and (for later purposes)

$$\Theta_{(-3)}^{\mu\nu} = (1/4\pi)(F_{(-1)}^{\mu\alpha}F_{\alpha(-2)}^\nu + F_{(-2)}^{\mu\alpha}F_{\alpha(-1)}^\nu + \tfrac{1}{2}g^{\mu\nu}F_{(-1)}^{\alpha\beta}F_{\alpha\beta(-2)}). \tag{III.4b}$$

Further, we find by a lengthy calculation, that[7a]

$$F_{(-1)}^{\alpha\beta}F_{\alpha\beta(-1)} = 0, \tag{III.5}$$

which then reduces $\Theta_{(-2)}^{\mu\nu}$ to the expression [using Eqs. (II.7)]

$$\Theta_{(-2)}^{\mu\nu} = \frac{1}{4\pi}\frac{R^\mu R^\nu}{\rho^6} \cdot \left\{ \left(\frac{ea^2}{2m} - \frac{e\mathring{a}_U}{2m} - \frac{3}{2}\frac{ea_U^2}{m}\right) \cdot \right.$$
$$\cdot \left[\left(\frac{ea^2}{2m} - \frac{e\mathring{a}_U}{2m} - \frac{3}{2}\frac{ea_U^2}{m}\right) M^{\alpha\beta}M_\beta{}^\delta R_\alpha R_\delta \right.$$
$$\left. - \frac{6ea_U}{2m}M^{\alpha\beta}\mathring{M}_\beta{}^\delta R_\alpha R_\delta - \frac{e}{m}M^{\alpha\beta}\ddot{M}_\beta{}^\delta R_\alpha R_\delta \right]$$
$$+ \frac{9e^2a_U^2}{4m^2}\mathring{M}^{\alpha\beta}\mathring{M}_\beta{}^\delta R_\alpha R_\delta + \frac{3e^2a_U}{2m^2}\mathring{M}^{\alpha\beta}\ddot{M}_\beta{}^\delta R_\alpha R_\delta$$
$$+ \frac{e^2}{4m^2}\ddot{M}^{\alpha\beta}\ddot{M}_\beta{}^\delta R_\alpha R_\delta - \rho^2\left(\frac{e^2a^2}{m} - \frac{e^2\mathring{a}_U}{m}\right)M^{\alpha\beta}U_\alpha a_\beta$$
$$+ \frac{3\rho^2e^2a_U}{m}\mathring{M}^{\alpha\beta}U_\alpha a_\beta + \frac{\rho e^2}{2m}\ddot{M}^{\alpha\beta}R_\alpha a_\beta$$
$$\left. + \frac{\rho e^2a_U}{m}\ddot{M}^{\alpha\beta}R_\alpha V_\beta + e^2\rho^2(a_U^2 - a^2) \right\}. \tag{III.6}$$

We note here that $\Theta_{(-2)}^{\mu\nu}$ consists of a scalar quantity, call it Λ, times $R^\mu R^\nu$.

Returning to the contribution that $\Theta_{(-2)}^{\mu\nu}$ makes in Eq. (III.3), we see that it is zero, as $\Lambda R^\mu R^\nu R_\nu = 0$. Thus, Eq. (III.3) has been verified.

The angular case requires more extensive considerations, as follows.

We note that the expression

$$\lim_{\rho \to \infty} \int J^{\alpha\mu\nu}R_\alpha d^2\omega = \lim_{\rho \to \infty} \int (\Theta^{\alpha\nu}X^\mu - \Theta^{\alpha\mu}X^\nu)R_\alpha d^2\omega$$

may only have nonzero contributions from the portions

1497   J. Math. Phys., Vol. 17, No. 8, August 1976

J. Cohn and H. Wiebe   1497
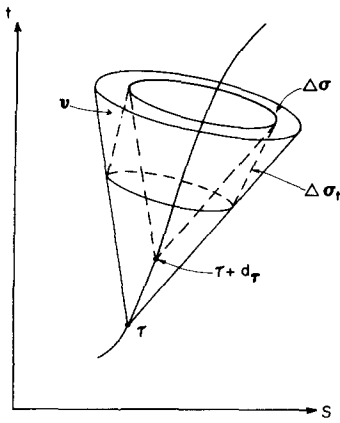
FIG. 2. Asymptotic space-like and timelike surfaces involved in momentum emission.

of $\Theta^{\lambda\sigma}$ depending on ρ^{-2} and ρ^{-3}. Again, as before, $\Theta^{\alpha\nu}_{(-2)}$ makes no contribution, so that we only need consider $\Theta^{\alpha\beta}_{(-3)}$.

Now, it can be shown by direct calculation that

$$F^{\alpha\beta}_{(-1)} F_{\alpha\beta(-2)} = 0 \tag{III. 7}$$

so that $\Theta^{\mu\nu}_{(-3)}$ has the form

$$\Theta^{\mu\nu}_{(-3)} = (1/4\pi) \cdot (F^{\mu\alpha}_{(-1)} F_\alpha{}^\nu_{(-2)} + F^{\mu\alpha}_{(-2)} F_\alpha{}^\nu_{(-1)}). \tag{III. 8}$$

Now, recalling the statement at the end of Sec. II, that $F^{\mu\nu}_{(-1)} R_\nu = 0$ and $F^{\mu\nu}_{(-2)} R_\nu = AR^\mu$, we see that

$$\Theta^{\alpha\nu}_{(-3)} R_\alpha = (1/4\pi) \cdot (F^{\alpha\beta}_{(-1)} R_\alpha F_\beta{}^\nu_{(-2)} - AR^\beta F_\beta{}^\nu_{(-1)}) = 0. \tag{III. 9}$$

Therefore,

$$\lim_{\rho \to \infty} \int J^{\alpha\mu\nu} R_\nu d^2\omega = 0 \tag{III. 10}$$

and our definition of angular momentum emission is also independent of the orientation of $(\Delta\sigma)$ as it should be.

In the next section we shall be interested in actually evaluating the expressions for the linear and angular momentum emission rates. To this end, we note here that the expression for the momenta [Eqs. (III. 1) and (III. 2)] can be altered so that they are easier to evaluate. Referring to Fig. 2, we see that by applying Gauss' law to the space–time volume v, and remembering the above result just proved, the integral over the annular region $(\Delta\sigma)$ can be replaced by an integration over the timelike strip $\Delta\sigma_t$, which is constructed to be parallel to V^μ_{ret}. This surface then has the components, $d\sigma^\mu = \rho^2 U^\mu d\Omega d\tau$, where $d\Omega$ denotes the differential of solid angle in 3-space. The expressions for momentum emission then become

$$dP^\mu_{\text{rad}} = -\lim_{\rho \to \infty} \int \Theta^{\mu\nu} U_\nu \rho^2 d\Omega d\tau \tag{III. 11}$$

and

$$dJ^{\mu\nu}_{\text{rad}} = -\lim_{\rho \to \infty} \int J^{\alpha\mu\nu} U_\alpha \rho^2 d\Omega d\tau. \tag{III. 12}$$

Finally, we see from these expressions that dP^μ_{rad} will only depend on $\Theta^{\mu\nu}_{(-2)}$, and $dJ^{\mu\nu}_{\text{rad}}$ will depend on $\Theta^{\mu\nu}_{(-3)}$ and possibly on $\Theta^{\mu\nu}_{(-2)}$.

IV. ASYMPTOTIC RADIATED MOMENTA

The asymptotic radiation rates for linear and angular momenta can now be calculated. However, since the general expressions turn out to be terribly lengthy, we shall confine ourselves to the "small spin" case, which means that we shall neglect all terms quadratic in the $M^{\alpha\beta}$, in the final expressions.

We proceed with the linear momentum case first:

Beginning with Eq. (III. 8a)

$$dP^\mu_{\text{rad}} = -\lim_{\rho \to \infty} \int \Theta^{\mu\nu} U_\nu \rho^2 d\Omega d\tau \tag{III. 11}$$

and recalling that only $\Theta^{\mu\nu}_{(-2)}$ contributes here, where $\Theta^{\mu\nu}_{(-2)}$ has the form $\Lambda R^\mu R^\nu$, we obtain

$$\frac{dP^\mu_{\text{rad}}}{d\tau} = -\lim_{\rho \to \infty} \int \Lambda R^\mu \rho^3 d\Omega. \tag{IV. 1}$$

Taking the expression for Λ from Eq. (III. 6) then gives the relation

$$\frac{dP^\mu_{\text{rad}}}{d\tau} = -\lim_{\rho \to \infty} \frac{1}{4\pi\rho^3} \int \left\{ \left(\frac{ea^2}{2m} - \frac{e\dot{a}_U}{2m} - \frac{3ea_U^2}{2m} \right) \right.$$
$$\cdot \left[\left(\frac{ea^2}{2m} - \frac{e\dot{a}_U}{2m} - \frac{3ea_U^2}{2m} \right) M^{\alpha\beta} M_\beta{}^\delta R_\alpha R_\delta \right.$$
$$\left. - \frac{3ea_U}{m} M^{\alpha\beta} \dot{M}_\beta{}^\delta R_\alpha R_\delta - \frac{e}{m} M^{\alpha\beta} \ddot{M}_\beta{}^\delta R_\alpha R_\delta \right]$$
$$+ \frac{9e^2 a_U^2}{4m^2} \dot{M}^{\alpha\beta} \dot{M}_\beta{}^\delta R_\alpha R_\delta + \frac{3e^2 a_U}{2m^2} \dot{M}^{\alpha\beta} \ddot{M}_\beta{}^\delta R_\alpha R_\delta$$
$$+ \frac{e^2}{4m^2} \ddot{M}^{\alpha\beta} \ddot{M}_\beta{}^\delta R_\alpha R_\delta - \rho^2 \left(\frac{e^2 a^2}{m} - \frac{e^2 \dot{a}_U}{m} \right) M^{\alpha\beta} U_\alpha a_\beta$$
$$+ \frac{3\rho^2 e^2 a_U}{m} M^{\alpha\beta} U_\alpha a_\beta + \frac{\rho e^2}{m} \ddot{M}^{\alpha\beta} R_\alpha a_\beta$$
$$\left. + \frac{\rho e^2 a_U}{m} \ddot{M}^{\alpha\beta} R_\alpha V_\beta + e^2 \rho^2 (a_U^2 - a^2) \right\} R^\mu d\Omega. \tag{IV. 2}$$

In order to carry out these integrations we must use the relation $R^\mu = \rho(U^\mu + V^\mu)$. Inserting this relation into the above equation and utilizing the well-known expression for the integral of a product of unit spacelike 4-vectors U^α gives a terribly lengthy expression, which is shortened by neglecting quadratic spin terms to the form

$$\frac{dP^\mu_{\text{s.s.}}}{d\tau} = \frac{e^2}{3} \left(\frac{a^2}{m} M^{\mu\beta} a_\beta - \frac{1}{m} \ddot{M}^{\mu\beta} a_\beta \right.$$
$$\left. - \frac{2}{m} \ddot{M}^{\alpha\beta} V_\alpha a_\beta V^\mu + 2a^2 V^\mu \right),$$

where the s. s. subscript signifies the small spin approximation made.

This is then the linear momentum emission rate in the small spin limit.

We next consider the angular momentum case: We start with the expression

$$dJ^{\mu\nu}_{\text{rad}} = -\lim_{\rho \to \infty} \int J^{\alpha\mu\nu} U_\alpha \rho^2 d\Omega d\tau \tag{III. 12}$$

and recall that $\Theta^{\mu\nu}_{(-2)}$ and $\Theta^{\mu\nu}_{(-3)}$ may contribute here, to

give

$$\frac{dJ^{\mu\nu}_{\text{rad}}}{d\tau} = -\lim_{\rho\to\infty}\int \left[(\Theta^{\alpha\nu}_{(-2)}X^{\mu} - \Theta^{\alpha\mu}_{(-2)}X^{\nu})\right.$$

$$\left. + (\Theta^{\alpha\mu}_{(-3)}X^{\mu} - \Theta^{\alpha\mu}_{(-3)}X^{\nu})\right]d\tilde\sigma_{\alpha}, \qquad \text{(IV.4)}$$

where $d\tilde\sigma_{\alpha} \equiv U_{\alpha}\rho^2\,d\Omega$.

The first integral above varies as ρ and the second as ρ^0, and we shall indicate the corresponding contributions to the total angular momentum emission with subscripts 1 and 0, respectively.

Now, using the fact that $R^{\mu} = X^{\mu} - Z^{\mu}$, we have

$$\frac{dJ^{\mu\nu}_{(1)}}{d\tau} = -\lim_{\rho\to\infty}\int \left[(\Theta^{\alpha\nu}_{(-2)}R^{\mu} - \Theta^{\alpha\mu}_{(-2)}R^{\nu})\right.$$

$$\left. + (\Theta^{\alpha\nu}_{(-2)}Z^{\mu} - \Theta^{\alpha\mu}_{(-2)}Z^{\nu})\right]d\tilde\sigma_{\alpha}. \qquad \text{(IV.5)}$$

Again, we recall that $\Theta^{\alpha\beta}_{(-2)} = \Lambda R^{\alpha}R^{\beta}$, so that the first bracketed term above gives no contribution. The surviving term which now varies as ρ^0 yields a very lengthy expression, which we avoid by taking the angular momentum about the event, Z^{μ}, of the retarded charge (i.e., we choose our origin at this event so that we may place $Z^{\mu} = 0$, for all μ). With this proviso then, we have that

$$\frac{dJ^{\mu\nu}_{(1)}}{d\tau} = 0. \qquad \text{(IV.6)}$$

Later, we shall indicate another way of expressing this term (in terms of the linear momentum) in the case that $Z^{\mu} \neq 0$.

We now have to evaluate

$$\frac{dJ^{\mu\nu}_{\text{rad}}}{d\tau} = \frac{dJ^{\mu\nu}_{(0)}}{d\tau} = -\lim_{\rho\to\infty}\int (\Theta^{\alpha\nu}_{(-3)}X^{\mu} - \Theta^{\alpha\mu}_{(-3)}X^{\nu})d\tilde\sigma_{\alpha}. \qquad \text{(IV.7)}$$

In this expression, we may replace X^{μ} and X^{ν} by R^{μ} and R^{ν}, since the contribution coming from Z^{μ} is zero in the limit as $\rho\to\infty$ (and also because in our special case we are taking $Z^{\mu} = 0$). Further, $\Theta^{\mu\nu}_{(-3)}$ may be taken as just $(1/4\pi)(F^{\mu\alpha}_{(-1)}F_{\alpha(-2)}^{\nu} + F^{\mu\alpha}_{(-2)}F_{\alpha(-1)}^{\nu})$, since $F^{\alpha\beta}_{(-1)}F_{\alpha\beta(-2)} = 0$. Then, in the small spin case we have

$$\Theta^{\mu\nu}_{\text{s.s.}(-3)} = \frac{1}{4\pi}\cdot\left\{R^{\mu}R^{\nu}\left(\frac{e^2}{m\rho^5}\ddot{M}^{\delta\alpha}U_{\delta}R_{\alpha} - \frac{e^2}{2m\rho^5}\dot{M}^{\delta\alpha}a_{\delta}U_{\alpha}\right.\right.$$

$$\left. + \frac{e^2 a_U}{m\rho^5}M^{\delta\alpha}a_{\delta}U_{\alpha}\right) - \frac{e^2 a_U}{2m\rho^5}(\dot{M}^{\mu\alpha}R_{\alpha}R^{\nu} + \dot{M}^{\nu\alpha}R_{\alpha}R^{\mu})$$

$$+ (\dot{M}^{\mu\alpha}R_{\alpha}R^{\nu} + \dot{M}^{\nu\alpha}R_{\alpha}R^{\mu})$$

$$\times\left[\frac{e}{\rho^4}\left(\frac{ea^2}{2m} - \frac{e\dot{a}_U}{2m} - \frac{3ea_U^2}{2m}\right) + \frac{3e^2 a_U^2}{2m\rho^4} - \frac{e^2 a^2}{2m\rho^4}\right]$$

$$+ (R^{\mu}U^{\nu} + R^{\nu}U^{\mu})\left(\frac{e^2 a_U}{2m\rho^4}\right)M^{\delta\alpha}a_{\alpha}U_{\delta}$$

$$- \frac{e^2}{m\rho^4}\dot{M}^{\delta\alpha}a_{\delta}U_{\alpha} + \frac{e^2 a_U}{\rho^4}\right) + (R^{\mu}V^{\nu} + R^{\nu}V^{\mu})$$

$$\cdot \frac{e^2 a_U}{2m\rho^4}M^{\delta b}U_{\delta}a_{\delta} - \frac{e^2}{2m\rho^5}(\ddot{M}^{\mu\alpha}R_{\alpha}R^{\nu} + \ddot{M}^{\nu\alpha}R_{\alpha}R^{\mu})$$

$$- \frac{e^2}{m\rho^4}\cdot(\dot{M}^{\mu\delta}a_{\delta}R^{\nu} + \dot{M}^{\nu\delta}a_{\delta}R^{\mu})\right\}.$$

Inserting this expression into Eq. (IV.7), with X^{λ} replaced by R^{λ}, and then integrating, then yields for the

angular momentum emission rate in the small spin limit the expression

$$\frac{dJ^{\mu\nu}_{\text{s.s.}}}{d\tau} = -\frac{e^2}{2}\left(-\frac{1}{m}\dot{M}^{[\nu\alpha}a_{\alpha}V^{\mu]} - \frac{1}{m}\ddot{M}^{[\nu\alpha}V_{\alpha}V^{\mu]}\right.$$

$$\left. + \frac{1}{3m}M^{[\nu\alpha}a_{\alpha}a^{\mu]} - \frac{4}{3}a^{[\nu}V^{\mu]} + \frac{2}{3m}\ddot{M}^{\mu\nu}\right), \qquad \text{(IV.9)}$$

where we have taken $Z^{\mu} = 0$, for all μ.

Before concluding this section, we now indicate how the requirement, $Z^{\mu} = 0$, may be relaxed in the above expression for angular momentum. Under a coordinate translation let the origin go into Z^{μ} ($\neq 0$). Also, let the subscripts i and f refer to the initial and final (translated) frames, respectively.

Then

$$X^{\mu}_f = X^{\mu}_i - Z^{\mu} \qquad \text{(IV.10)}$$

so that

$$\left(\frac{dJ^{\mu\nu}}{d\tau}\right)_f = \left(\frac{dJ^{\mu\nu}}{d\tau}\right)_i$$

$$- \lim_{\rho\to\infty}\int (\Theta^{\alpha\nu}Z^{\mu} - \Theta^{\alpha\mu}Z^{\nu})d\tilde\sigma_{\alpha}. \qquad \text{(IV.11)}$$

The Z^{α} are constant in these integrations, so we obtain

$$\left(\frac{dJ^{\mu\nu}}{d\tau}\right)_f = \left(\frac{dJ^{\mu\nu}}{d\tau}\right)_i + \left(Z^{\mu}\frac{dP^{\nu}}{d\tau} - Z^{\nu}\frac{dP^{\mu}}{d\tau}\right). \qquad \text{(IV.12)}$$

Thus, using Eqs. (IV.3) and (IV.9), we can find the angular momentum emission rate about any origin, for small spin.[7b]

V. IMPLICATIONS

In this section we consider two implications of the preceding discussion: the center-of-energy theorem for accelerating charges, and the lack of invariance of the energy emission rate for spinning charges, respectively.

The fact that a particle may emit angular momentum even if it is not spinning (or moving along a curve) leads to a theorem concerning the location of the "center of radiated energy" of an arbitrarily moving nonspinning charge.[8]

The center of radiated energy in the retarded rest frame of a nonspinning charge is defined by the relation

$$R^i_c \equiv \lim_{\rho\to\infty}\int_{(\Delta\sigma)} X^i\,\Theta^{44}d^3X / \int_{(\Delta\sigma)}\Theta^{44}d^3X, \qquad \text{(V.1)}$$

where $(\Delta\sigma)$ denotes the usual spacelike annular hypersurface, $d^3X = -d\sigma_4$, and all quantities are evaluated in the rest frame of the retarded charge.

Now, the above expression in the denominator can be expressed as

$$\lim_{\rho\to\infty}\int \Theta^{44}d^3X = -dW_{\text{r.r.}}, \qquad \text{(V.2)}$$

where $dW_{\text{r.r.}}$ is the energy radiated by the charge (in its retarded rest frame) during the proper time $d\tau$, and the above expression in the numerator appears as part of

the expression for dJ^{4i} for a nonspinning charge—i. e.,

$$dJ^{4i}_{r.t.} (\text{no spin}) = \lim_{\rho \to \infty} \int_{(\Delta\sigma)} (\Theta^{44}_{n.s.} X^i - \Theta^{4i}_{n.s.} X^4) d^3 X. \quad (V.3)$$

The second integral above is given by

$$\lim_{\rho \to \infty} \int_{(\Delta\sigma)} \Theta^{4i} d^3 X = - dP^i_{r.t.} = 0 \quad (V.4)$$

since the 3-momentum emission in the retarded rest frame of the charge is zero [see Eq. (IV.3)].

Thus,

$$\lim_{\rho \to \infty} \int_{(\Delta\sigma)} \Theta^{44} X^i d^3 X = dJ^{4i}_{r.t.} (\text{no spin}). \quad (V.5)$$

Using these results in Eq. (V.1), we now have

$$R^i_c = - \left.\frac{dJ^{4i}}{dW}\right)_{r.t.} = - \frac{dJ^{4i}}{d\tau} \circ \left.\left[\frac{dW}{d\tau}\right]^{-1}\right)_{r.t.} \quad (V.6)$$

Now, $(dW/d\tau)_{r.t.}$ is merely the 4th component of radiated linear momentum in the rest frame of the nonspinning charge and is given by Eq. (IV.3) as

$$\left(\frac{dW}{d\tau}\right)^{n.s.}_{r.t.} = \tfrac{2}{3} e^2 a^2. \quad (V.7)$$

Further, $(dJ^{4i}/d\tau)^{n.s.}_{r.t.}$ is given by Eqs. (IV.10) and (IV.13) as

$$\left(\frac{dJ^{4i}}{d\tau}\right)^{n.s.}_{r.t.} = \tfrac{2}{3} e^2 (a^i - a^2 Z^i)_{r.t.} \quad (V.8)$$

Therefore, we finally have

$$R^i_c = \{- (1/a^2)(a^i - a^2 Z^i)\}_{r.t.} \quad (V.9)$$

so that the center of radiated energy only depends on the particle's retarded acceleration and location and is therefore *constant*, unlike the center-of-energy theorem for free radiation fields where the *velocity* of the center of energy is constant.[9]

Finally, we consider the lack of invariance of the energy emission rate for a spinning charge.

Now, for nonspinning charges, one has for the energy—momentum emission rate the expression

$$\left(\frac{dP^{\mu}_{rad}}{d\tau}\right)_{n.s.} = \tfrac{2}{3} e^2 a^2 V^{\mu} \quad (V.10)$$

from which one easily shows[10] that dW/dt is invariant.

We will now show that dW/dt is *not* invariant for *spinning* charges.[11]

We return, then, to Eq. (IV.3), which gives

$$\left(\frac{dW}{d\tau}\right)_{s.s.} = \left(\frac{dP^4}{d\tau}\right)_{s.s.}$$

$$= \frac{e^2}{3}\left(\frac{a^2}{m} M^{4\beta} a_\beta - \frac{1}{m}\ddot{M}^{4\beta} a_\beta\right.$$

$$\left. - \frac{2}{m}\ddot{M}^{\alpha\beta} V_\alpha a_\beta V^4 + 2a^2 V^4\right).$$

And this gives then

$$\left(\frac{dW}{dt}\right)_{s.s.} = \frac{e^2}{3}\frac{1}{\gamma}$$

$$\circ \left(\frac{a^2}{m} M^{4\beta} a_\beta - \frac{1}{m}\ddot{M}^{4\beta} a_\beta - \frac{2}{m}\ddot{M}^{\alpha\beta} V_\alpha a_\beta \gamma + 2a^2\gamma\right).$$

$$(V.12)$$

Now, the last two terms on the right-hand side here comprise an invariant (when multiplied through by the factor in front of the bracket), but the first two terms do not—and this completes the statement.

To put this proof into perspective, however, we must say a few words about the procedure of neglecting the quadratic spin terms [Eq. (IV.3)] in its execution. Neglect of these terms is tantamount to assuming that, for given values of the linear kinematical quantities, spin terms like $M^{\alpha\beta}$ or $\dot{M}^{\alpha\beta}$ can be made arbitrarily small. We assume this could be done with suitable (perhaps nonelectromagnetic) forces. That is, in the present development, we are not concerned with the existence of equations of motion which would *determine* quantities like $\dot{M}^{\alpha\beta}$ or $\ddot{M}^{\alpha\beta}$ in terms of linear kinematical quantities. Thus, we must concede that it might result that the proper equations of motion of an electron are such that the above proof does not apply.

---

*This work is part of a Ph. D. dissertation submitted by H.W.

[†]Now at the Department of Mathematics, Wesleyan University, Delaware, Ohio.

[1]J. Cohn, J. Math. Phys. 10, 799 (1969).

[2]M. Kolsrud and E. Leer, Physica Norvegica 2, 181 (1967).

[3]F. Rohrlich, Classical Charged Particles (Addison-Wesley, Reading, Mass., 1965), Chap. IV.

[4]J. N. Goldberg and R. P. Kerr, J. Math. Phys. 5, 172 (1964).

[5]See Ref. 1.

[6]See Ref. 3, p. 282. However, for a more complete explication of this notion, see J. L. Synge, Ann. di Mat. 84, 33 (1970). The author is indebted to the referee for drawing his attention to this reference and for making some related clarifying comments.

[7a]This result is also found by Kolsrud and Leer, see Ref. 2.

[7b]For a different approach giving an approximate expression for the linear and angular momentum emission rates, see Van Weert, Physica 66, 79 (1973).

[8]This theorem is a variation of the center of energy theorem for free radiation fields; see Ref. 3, Chap. IV.

[9]This theorem was first discussed in Ref. 1. In that discussion, however, a mistake was made in a sign, and a factor 2 was omitted, and the origin was taken at the retarded charge event.

[10]See Ref. 3, Chap. V.

[11]It seems to be commonly accepted (see Ref. 2, for example) that the emission rate of energy from any finite distribution of charge must be an ivariant, as it is for nonspinning charges. However, the author knows of no rigorous proof of such a statement. A "proof" is given in J. D. Jackson, Classical Electrodynamics (Wiley, New York, 1962), p. 469— where the statement is made that, since the energy emitted by a moving charge transforms like time (i.e., like the fourth component of a contravariant 4-vector) their quotient must be an invariant. But this statement need not be true— viz., the quotient of the fourth components of 4-acceleration and 4-velocity is certainly not invariant.

Exact closed evolution equation for the electron density operator averaged over impurity configurations

Shigeji Fujita

Departamento de Física y Química, Universidad Autonoma Metropolitana/Iztapalapa, México 13, D. F., Mexico
and Department of Physics and Astronomy, State University of New York at Buffalo, Buffalo, New York 14214*
(Received 15 September 1975)

The exact Markoffian evolution equation for the electron density operator $n(t)$ averaged over impurity configurations (describing a grand ensemble of non-interacting electrons in the potential fields of impurities fixed in space) is not closed. The interaction term, that is, the integral involving the electron-impurity interaction and the electron-impurity density operator is analyzed with the aid of diagrams after its evolution operator is expanded in perturbation and its initial density operators are expanded in terms of correlation operators. A closed non-Markoffian equation for $n(t)$ is obtained without introducing any approximations. This equation contains infinite sets of collision and initial correlation terms which are most conveniently represented by connected diagrams, and which can be expressed in terms of $n(t)$ and initial correlation operators arbitrarily given. The equation is found to be identical with the closed evolution equation obtained earlier in the fixed-particle-number ensemble theory with the bulk limit.

1. INTRODUCTION

A system of noninteracting electrons in a potential field of impurities fixed in space, is a simple model for a conducting solid at very low temperatures, where electrons are mainly scattered by impurities. In discussing various transport phenomena, the Boltzmann equation can provide a good starting point.[1] This equation is an approximate closed equation for the single-electron density operator $n(t)$ averaged over the impurity configurations. The exact Markoffian equation (hierarchy equation) for $n(t)$, see Eq. (3.24) [which may be obtained by averaging the Liouville equation over impurity configurations] is known to contain another unknown, the electron-impurity density operator, and thus this equation is not a closed equation. In the past, many attempts have been made to derive and generalize the Boltzmann equation by introducing approximations at various stages; many interesting results are known.[2] A fundamental question however is: Can the Markoffian hierarchy equation be transformed into a closed equation without introducing approximations? The present paper deals with this question and gives a definite affirmative answer.

In earlier works,[3,4] the present author showed that the hierarchy equation, (3.24), can be brought into a closed non-Markoffian equation,[3] see (5.9.27), rigorously in the *bulk limit*: N (number of electrons) $\to \infty$, N_s (number of impurities) $\to \infty$, Ω (volume) $\to \infty$, while N/Ω and N_s/Ω remain finite. This was established in the theory dealing with an ensemble of systems with fixed numbers of particles (electrons or impurities), called a *fixed-particle-number ensemble* hereafter. In the course of the analysis quantities like

$$\int d^3R_2 f_{N_s}^{(2)}(\mathbf{R}_1, \mathbf{R}_2) - [\int d^3R_2 f_{N_s}^{(1)}(\mathbf{R}_2)] f_{N_s}^{(1)}(\mathbf{R}_1)$$

$$= [N_s - 1] f_{N_s}^{(1)}(\mathbf{R}_1) - N_s f_{N_s}^{(1)}(\mathbf{R}_1)$$

$$= -f_{N_s}^{(1)}(\mathbf{R}_1),$$

where $f_{N_s}^{(1)}$ and $f_{N_s}^{(2)}$ are one-impurity and two-impurity

distribution functions defined in the N_s-impurity ensemble, were dropped against the individual integrals; while this is a neglect of a quantity of the order $1/N_s$, the accumulation of infinitely many such terms could contribute a finite amount even in the bulk limit; this may possibly destroy the final result.

In order to overcome this difficulty, we will present a new theory, utilizing a grand ensemble of systems with variable numbers of particles. We will show that the hierarchy equation, (3.24), for the grand-ensemble-averaged electron density operator can rigorously be transformed into a closed equation, (7.9), without invoking the bulk limit. The resulting equation is formally identical with the result obtained in the fixed-particle-number ensemble theory with the bulk limit. Thus, the present theory not only removes any doubt about the previous result but also establishes an exact transformation. The latter fact immediately implies that the closed equation accords a reversible motion just as the original hierarchy equation. This is a significant conclusion since the bulk limit which was used in conjunction with the fixed-particle-number ensemble theory has clouded the question of the reversibility with respect to the said closed evolution equation. The present theory is developed for a system with arbitrary impurity configurations characterized by many-impurity distribution functions; the usual assumption of the uniform distribution can be obtained as a special case.

The theoretical development of the present paper partially parallels that of the earlier work[3] dealing with the same problem in the framework of the fixed-particle-number ensemble. In particular, the expansion of the evolution operator into a perturbation series is the same. The specification of the initial condition in terms of grand-ensemble correlation operators is newly prescribed, however. For part of this development, the author draws materials from the book by Yvon.[5] The diagram representation and analysis proceed in a parallel manner except that Theorem C, which concerns a special class of non-contributing connected diagrams, is used in the grand ensemble theory.

In Sec. 2, the Liouville equations for many-electron and single-electron density operators are discussed. In Sec. 3 a grand ensemble of systems with variable numbers of particles is introduced. Correlation operators are introduced in Sec. 4. In Sec. 5 the reduced density operator averaged over the impurity configurations is expanded, and terms in the expansion are represented by diagrams. These diagrams are analyzed in Sec. 6. In Sec. 7, the interaction term which appears in the hierarchy equation, is analyzed in a similar manner, and a closed evolution equation is obtained. Remarks on the nature of the closed equation are given in Sec. 8.

Throughout the text the units are chosen such that $\hbar = 1$.

2. THE LIOUVILLE EQUATION

Let us consider a system of free electrons in interaction with static impurities, characterized by the Hamiltonian

$$H = \sum_{j=1}^{N} h_0^{(j)} + \lambda \sum_{j=1}^{N} \sum_{\alpha=1}^{N_s} \bar{v}_\alpha^{(j)} \tag{2.1}$$

$$\equiv \sum_j (h_0^{(j)} + \lambda v^{(j)}) \equiv \sum_j h^{(j)}, \tag{2.2}$$

$$h_0^{(j)} \equiv h_0(\mathbf{r}_j \mathbf{p}_j) \equiv \frac{1}{2m} [\mathbf{p}_j + e\mathbf{A}(\mathbf{r}_j)]^2 - e\phi(\mathbf{r}_j), \tag{2.3}$$

$$\bar{v}_\alpha^{(j)} \equiv \bar{v}(\mathbf{r}_j - \mathbf{R}_\alpha), \tag{2.4}$$

where m and $-e$ are respectively the mass and charge of a spinless electron; Φ and A are scalar and vector potentials.

We assume that the impurities are fixed in space, and we thus neglected their kinetic energies in our Hamiltonian H in (2.1). The impurity configuration may be specified by giving the positions of N_s impurities $\mathbf{R}_1$, $\mathbf{R}_2, \ldots, \mathbf{R}_{N_s}$ in a probabilistic manner.

The density matrix in position space for the total system is denoted by

$$\rho(\mathbf{r}_1 \mathbf{r}_2 \cdots \mathbf{r}_N; \mathbf{r}_1' \mathbf{r}_2' \cdots \mathbf{r}_N'; \mathbf{R}_1 \mathbf{R}_2 \cdots \mathbf{R}_{N_s}; t). \tag{2.4}$$

The (density) operator corresponding to this matrix changes with time according to the Liouville equation

$$i \frac{\partial \rho(t)}{\partial t} = [H, \rho(t)] \equiv H\rho(t) - \rho(t)H. \tag{2.5}$$

Since our Hamiltonian H is the sum of single-electron energies, no correlation is expected between electrons except that coming from the (Fermi) statistical origin; in the case of a grand canonical ensemble the latter can simply be handled by the Fermi operator technique. Also see below. The correlation between impurities [which remains constant in time if ever present] can be handled in a standard manner. If no correlation and the invariance under translation are assumed, the impurity configuration is specified by a single constant, i.e., the impurity density $n_s \equiv N_s/\Omega$. This is the case for which the great majority of works have been done.[2] The correlation between electrons and impurities can, however, be created in time by the interaction $\bar{v}_\alpha^{(j)}$, and must be treated with care.

In second quantization, the Hamiltonian corresponding to (2.1) can be written as

$$H = \int d^3 r_1 \, d^3 r_2 \, \langle \mathbf{r}_1 | h | \mathbf{r}_2 \rangle \psi^\dagger(\mathbf{r}_1) \psi(\mathbf{r}_2), \tag{2.6}$$

where ψ and $\psi^\dagger$ are annihilation and creation operators in position and satisfy the following anticommutation rules:

$$\{\psi(\mathbf{r}), \psi^\dagger(\mathbf{r}')\} \equiv \psi(\mathbf{r})\psi^\dagger(\mathbf{r}') + \psi^\dagger(\mathbf{r}')\psi(\mathbf{r}) = \delta^{(3)}(\mathbf{r} - \mathbf{r}'),$$

$$\{\psi(\mathbf{r}), \psi(\mathbf{r}')\} = \{\psi^\dagger(\mathbf{r}), \psi^\dagger(\mathbf{r}')\} = 0; \tag{2.7}$$

$\delta^{(3)}$ denotes the three-dimensional Dirac delta function. The position density matrix is defined by

$$n(\mathbf{r}_1, \mathbf{r}_2; \mathbf{R}_1 \mathbf{R}_2 \cdots \mathbf{R}_{N_s}, t) \equiv \text{Tr}\{\psi^\dagger(\mathbf{r}_2)\psi(\mathbf{r}_1)\rho(t)\} \equiv \langle \mathbf{r}_1 | n(t) | \mathbf{r}_2 \rangle, \tag{2.8}$$

where the symbol Tr means the trace with respect to the N-electron states. It was shown in our earlier work[3] that the single-electron density operator $n(t)$ defined through (2.8), which in general depends on the impurity configuration, obeys the following equation:

$$i \frac{\partial n}{\partial t} = [h, n(t)] = [h_0, n(t)] + \lambda \sum_\alpha [\bar{v}_\alpha, n(t)]. \tag{2.9}$$

In the present paper, we will be interested in the behavior of the (electron) current density only. The latter can be obtained through the single-electron density operator $\bar{n}(t)$ with the impurity-configuration-average completed:

$$\overline{n(t)} \equiv \langle n(t) \rangle_{\text{impurity average}}. \tag{2.10}$$

The $\overline{n(t)}$ can be sought from Eq. (2.8). However, the impurity average of Eq. (2.8) does not give a closed equation for $\bar{n}(t)$ since the *interaction term*, that is, the second term of the third member gives rise to

$$\lambda \left\langle \sum_\alpha [\bar{v}_\alpha, n(t)] \right\rangle, \tag{2.11}$$

which generates the dynamic correlation between electron and impurities.

3. THE GRAND ENSEMBLE

Since measurable thermodynamic quantities are the so-called intensive quantities, that is, densities of some sort, the precise numbers of electrons and impurities, N and N_s, do not enter into the final expressions for these quantities. We may well consider a grand ensemble of systems with variable numbers of particles.

Let us denote the probability that the system contains N electrons and N_s impurities by Π_{N, N_s}, where both N and N_s can take any nonnegative integers $0, 1, 2, \cdots$. The set of Π_{N, N_s} will be subject to

$$\sum_{N=0}^{\infty} \sum_{N_s=0}^{\infty} \Pi_{N, N_s} = 1. \tag{3.1}$$

The average numbers of particles contained in the volume can now be expressed by

$$\langle N \rangle = \sum_{N, N_s} N \Pi_{N, N_s}, \tag{3.2}$$

$$\langle N_s \rangle = \sum_{N, N_s} N_s \Pi_{N, N_s}. \tag{3.3}$$

When the system contains N electrons and N_s im-

purities, the corresponding density operator will be denoted by

$$\rho_{N,N_s} \equiv \rho_{N,N_s}(\mathbf{R}_1, \mathbf{R}_2, \ldots, \mathbf{R}_{N_s}). \tag{3.4}$$

with the subscripts denoting the N and N_s dependence. We may assume the following normalization condition:

$$\mathrm{Tr}'_{,N_s}\mathrm{Tr}_{,N}\{\rho_{N,N_s}\}$$
$$\equiv \frac{1}{N_s!}\int \cdots \int d^3R_1\, d^3R_2 \cdots d^3R_{N_s}\,\mathrm{Tr}_{,N}\{\rho_{N,N_s}\}$$
$$= \Pi_{N,N_s}. \tag{3.5}$$

We introduce the single-electron density matrix by way of Eq. (2.8):

$$n_{N,N_s}(\mathbf{r}_1, \mathbf{r}_2; \mathbf{R}_1, \mathbf{R}_2, \ldots, \mathbf{R}_{N_s}) \equiv \mathrm{Tr}_{,N}\{\psi^\dagger(\mathbf{r}_2)\psi(\mathbf{r}_1)\rho_{N,N_s}\}. \tag{3.6}$$

Summing over all N, we obtain

$$n_{N_s}(\mathbf{r}_1, \mathbf{r}_2; \mathbf{R}_1, \mathbf{R}_2, \ldots, \mathbf{R}_{N_s})$$
$$= \sum_{N=1}^{\infty} n_{N,N_s}(\mathbf{r}_1, \mathbf{r}_2; \mathbf{R}_1, \mathbf{R}_2, \ldots, \mathbf{R}_{N_s}). \tag{3.7}$$

This matrix depends on the electron variables, $(\mathbf{r}_1, \mathbf{r}_2)$, as well as on the positions of impurities, $(\mathbf{R}_1, \mathbf{R}_2, \ldots, \ldots, \mathbf{R}_{N_s})$. The corresponding density operator n_{N_s} obeys the Liouville equation of the form (2.8) just as the operator n_{N,N_s}. The diagonal sum of n_{N_s}, denoted by $\mathrm{tr}\{n_{N_s}\}$, gives the distribution function for N_s impurities:

$$\mathrm{tr}\{n_{N_s}\} \equiv f_{N_s}(\mathbf{R}_1, \mathbf{R}_2, \ldots, \mathbf{R}_{N_s}). \tag{3.8}$$

If we further take the "trace" with respect to the impurities [in the sense defined in Eqs. (3.5)], we obtain

$$\mathrm{Tr}'_{,N_s}\{f_{N_s}\} = \sum_{N=0}^{\infty} \Pi_{N,N_s} \equiv \Pi_{N_s}. \tag{3.9}$$

The Π_{N_s} can now be interpreted as the probability of having N_s impurities in the system.

Let us further introduce reduced quantities by

$$n_{N_s}^{(M)}(\mathbf{r}_1, \mathbf{r}_2; \mathbf{R}_1, \mathbf{R}_2, \ldots, \mathbf{R}_M)$$
$$\equiv \frac{1}{(N_s - M)!}\int \cdots \int d^3R_{M+1}\, d^3R_{M+2} \cdots d^3R_{N_s}$$
$$\times n_{N_s}(\mathbf{r}_1, \mathbf{r}_2; \mathbf{R}_1, \mathbf{R}_2, \ldots, \mathbf{R}_{N_s}),$$
$$M = 0, 1, 2, \ldots, N_s - 1. \tag{3.10}$$

For $M = 0$, $n_{N_s}^{(0)}(\mathbf{r}_1, \mathbf{r}_2) \equiv n_{N_s}$ gives the $\mathbf{r}_1 - \mathbf{r}_2$ element of the electron density matrix averaged over impurity configurations. More generally, $n^{(M)}$ corresponds to the reduced density matrix for one electron and M impurities. These quantities are interrelated as follows:

$$\int d^3R_1\, n_{N_s}^{(1)}(\mathbf{r}_1, \mathbf{r}_2; \mathbf{R}_1) = N_s n_{N_s}^{(0)}(\mathbf{r}_1, \mathbf{r}_2),$$
$$\int d^3R_M\, n_{N_s}^{(M)}(\circ\circ; \mathbf{R}_1, \mathbf{R}_2, \ldots, \mathbf{R}_M) \tag{3.11}$$
$$= (N_s - M + 1)n_{N_s}^{(M-1)}(\circ\circ; \mathbf{R}_1, \mathbf{R}_2, \ldots, \mathbf{R}_{M-1}).$$

In normal applications of the theory, the density matrices with small M, say, $M = 0, 1$, are needed. It is therefore tempting to approximate the factor $(N_s - M + 1)$ by N_s, which is quite large, that is, of the order 10^{20}. We, however, reject this temptation [as we explained the reason in the Introduction]; we develop a grand-ensemble theory and proceed in a rigorous manner.

Let us introduce reduced quantities for the grand ensemble by

$$n^{(M)} \equiv \sum_{N_s = M}^{\infty} n_{N_s}^{(M)}(\mathbf{R}_1, \ldots, \mathbf{R}_M),$$
$$f^{(M)} \equiv \mathrm{tr}\ \{n^{(M)}\} \equiv \int d^3r\, n^{(M)}(\mathbf{r}, \mathbf{r}; \mathbf{R}_1, \ldots, \mathbf{R}_M) \tag{3.12}$$
$$= \sum_{N_s = M}^{\infty} f_{N_s}^{(M)}(\mathbf{R}_1, \ldots, \mathbf{R}_M). \tag{3.13}$$

By simple calculation we can show the following relations

$$\langle N\rangle = \mathrm{tr}\{n^{(0)}\} \equiv \mathrm{tr}\{n\}, \tag{3.14}$$
$$\langle N_s\rangle = \int d^3R_1 f^{(1)}(\mathbf{R}_1) = \int d^3R_1\,\mathrm{tr}\{n^{(1)}(\mathbf{R}_1)\}. \tag{3.15}$$

We further note that there are no simple relations like (3.11) which as in the case where the number of impurities is fixed, allow us to pass from $n^{(M)}$ to $n^{(M-1)}$ by integration.

From (3.10) and (3.12), we can write a general expression for $n^{(M)}$ in terms of $\{n_{N_s}\}$ as follows:

$$n^{(M)}(\mathbf{R}_1, \mathbf{R}_2, \ldots, \mathbf{R}_M) \equiv \sum_{N_s \geqslant M}^{\infty} \frac{1}{(N_s - M)!}$$
$$\int \circ\circ\circ \int dR_{M+1}\, dR_{M+2} \circ\circ\circ dR_{N_s}\, n_{N_s}(\mathbf{R}_1, \mathbf{R}_2, \ldots, \mathbf{R}_{N_s}). \tag{3.16}$$

Likewise, n_{N_s} can be obtained in terms of $n^{(M)}$:

$$n_{N_s}(\mathbf{R}_1, \mathbf{R}_2, \ldots, \mathbf{R}_{N_s})$$
$$= \sum_{L=0}^{\infty} \frac{(-1)^L}{L!}\int \circ\circ\circ \int dR_{N_s+1}\, dR_{N_s+2} \circ\circ\circ dR_{N_s+L}$$
$$\times n^{(N+L)}(\mathbf{R}_1, \mathbf{R}_2, \ldots, \mathbf{R}_{N+L}). \tag{3.17}$$

The correspondence between $\{n^{(M)}\}$ and $\{n_{N_s}\}$ is therefore one-to-one.

A physical quantity A of the system depends in general on the dynamical variables describing electrons and on the configuration of impurities. Its average with the N-electron-and-N_s-impurity ensemble is given by

$$\langle A\rangle_{N,N_s} = \mathrm{Tr}_{,N}\{A_{N,N_s}\rho_{N,N_s}\} \times \Pi_{N,N_s}^{-1}. \tag{3.18}$$

Its average with the grand ensemble is defined by

$$\langle A\rangle \equiv \sum_{N=0}^{\infty}\sum_{N_s=0}^{\infty} \Pi_{N,N_s}\langle A\rangle_{N,N_s} = \sum_N\sum_{N_s}\mathrm{Tr}_{,N_s}\{A_{N,N_s}\rho_{N,N_s}\}$$
$$\equiv \mathrm{TR}\{A\rho\}, \tag{3.19}$$

where the new symbol, TR (grand trace), indicates the diagonal sum in the grand ensemble. In particular, if A is the sum of terms depending on the single-electron variables only, and is given by

$$A = \sum_j a^{(j)} = \int\int d^3r_1\, d^3r_2\,\langle \mathbf{r}_1|a|\mathbf{r}_2\rangle\psi^\dagger(\mathbf{r}_1)\psi(\mathbf{r}_2), \tag{3.20}$$

then its average with the grand ensemble is given by

$$\langle A\rangle = \int\int d^3r_1\, d^3r_2\,\langle \mathbf{r}_1|a|\mathbf{r}_2\rangle n(\mathbf{r}_2, \mathbf{r}_1) \equiv \mathrm{tr}\{an\}. \tag{3.21}$$

By a similar calculation one can show that the grand ensemble average of a double sum of the form

$$B \equiv \sum_j\sum_\alpha b_\alpha^{(j)} \equiv \int\int d^3r_1\, d^3r_2\sum_\alpha \langle \mathbf{r}_1|b(\mathbf{R}_\alpha)|\mathbf{r}_2\rangle\psi^\dagger(\mathbf{r}_1)\psi(\mathbf{r}_2) \tag{3.22}$$

can be expressed in terms of one-electron–one-impurity density matrix $n^{(1)}(\mathbf{R})$:

1503   J. Math. Phys., Vol. 17, No. 8, August 1976

Shigeji Fujita   1503

$$\langle B \rangle = \int d^3R \int\int d^3r_1 d^3r_2 \langle \mathbf{r}_1 | b(\mathbf{R}) | \mathbf{r}_2 \rangle n^{(1)}(\mathbf{r}_2, \mathbf{r}_1; \mathbf{R}). \quad (3.23)$$

In a system which neither gains nor loses particles both Π_{N,N_s} and Π_{N_s} are constants, and each ρ_{N,N_s} satisfies the Liouville equation (2.4). It follows that each one-electron density operator n_{N,N_s} satisfies Eq. (2.8). Integrating the latter with respect to $\mathbf{R}_1, \mathbf{R}_2, \ldots, \mathbf{R}_{N_s}$, multiplying by $(N_s!)^{-1}$, and subsequently summing with respect to N and N_s, we obtain

$$i\frac{\partial n(t)}{\partial t} = [h_0, n] + \lambda \int d^3R[\bar{v}(\mathbf{r} - \mathbf{R}), n^{(1)}(\mathbf{R}, t)]. \quad (3.24)$$

This equation explicitly shows that the evolution of the electron density operator $n(t)$ is connected with the one-electron—one-impurity density operator $n^{(1)}(\mathbf{R}, t)$. This equation is sometimes referred to as the *hierarchy equation* of the first degree. Since it contains two unknown n and $n^{(1)}$, it cannot be solved in the present form. We will later derive a closed equation for $n(t)$ from this equation.

4. CORRELATION OPERATORS

Since our system contains similar impurities, density operators $n_{N_s}(\mathbf{R}_1, \mathbf{R}_2, \ldots, \mathbf{R}_{N_s})$ must be symmetric, that is, invariant under the permutations of impurity indices. It follows that reduced density operators $n^{(M)}(\mathbf{R}_1, \mathbf{R}_2, \ldots, \mathbf{R}_M)$ are also symmetric. We will now introduce correlation operators which satisfy the same symmetry.

The correlation functions χ_l involving impurities only, are defined through the following equations:

$$f^{(1)}(\mathbf{R}_1) \equiv f(\mathbf{R}_1),$$

$$f^{(2)}(\mathbf{R}_1, \mathbf{R}_2) \equiv f(\mathbf{R}_1)f(\mathbf{R}_2) + \chi_2(\mathbf{R}_1, \mathbf{R}_2),$$

$$f^{(3)}(\mathbf{R}_1, \mathbf{R}_2, \mathbf{R}_3) \equiv f(\mathbf{R}_1)f(\mathbf{R}_2)f(\mathbf{R}_3) + f(\mathbf{R}_1)\chi_2(\mathbf{R}_2, \mathbf{R}_3)$$

$$+ f(\mathbf{R}_2)\chi_2(\mathbf{R}_1, \mathbf{R}_3) + f(\mathbf{R}_3)\chi_2(\mathbf{R}_1, \mathbf{R}_2) + \chi_3(\mathbf{R}_1, \mathbf{R}_2, \mathbf{R}_3),$$

$$\circ\circ\circ.$$

(4.1)

Note that the suffices on χ indicate the numbers of impurities. The correlation operators involving electron and impurities are defined as follows:

$$n^{(0)} \equiv n,$$

$$n^{(1)}(\mathbf{R}_1) \equiv f(\mathbf{R}_1)n + \nu_1(\mathbf{R}_1),$$

$$n^{(2)}(\mathbf{R}_1, \mathbf{R}_2) \equiv f(\mathbf{R}_1)f(\mathbf{R}_2)n + f(\mathbf{R}_1)\nu_1(\mathbf{R}_2) + f(\mathbf{R}_2)\nu_1(\mathbf{R}_1)$$

$$+ \chi_2(\mathbf{R}_1, \mathbf{R}_2)n + \nu_2(\mathbf{R}_1, \mathbf{R}_2), \quad (4.2)$$

$$\cdots.$$

By taking the electron trace of Eqs. (4.2), using Eqs. (3.13) and comparing the results with Eqs. (4.3), we obtain

$$\mathrm{tr}\{\nu^{(M)}(\mathbf{R}_1, \mathbf{R}_2, \ldots, \mathbf{R}_M)\} = 0, \quad M = 1, 2, \circ\circ\circ. \quad (4.3)$$

It is clear that correlation functions χ_l and correlation operators ν_k can be expressed in terms of $f^{(M)}$ and $n^{(M)}$, e.g.,

$$\chi_2(\mathbf{R}_1, \mathbf{R}_2) = f^{(2)}(\mathbf{R}_1, \mathbf{R}_2) - f(\mathbf{R}_1)f(\mathbf{R}_2),$$

$$\nu_1(\mathbf{R}_1) = n^{(1)}(\mathbf{R}_1) - f(\mathbf{R}_1)n. \quad (4.4)$$

The correspondence between correlation operator ν_k and reduced density operators $n^{(M)}$ is therefore unique.

Since the reduced operators $\{n^{(M)}\}$ are related to the density operators $\{n_{N_s}\}$ by Eqs. (3.16) correlation operators $\{\nu_k\}$ can be expressed in terms of $\{n_{N_s}\}$. The converse is also true. We can show (by the cumulant method, also see below) that

$$f_0 = \exp[-\int d^3R_1 f(\mathbf{R}_1) + \tfrac{1}{2}\int\int d^3R_1 d^3R_2\, \chi_2(\mathbf{R}_1, \mathbf{R}_2) - \cdots],$$

$$n_0/f_0 = n - \int d^3R_1\, \nu_1(\mathbf{R}_1) + \tfrac{1}{2}\int\int d^3R_1 d^3R_2\, \nu_2(\mathbf{R}_1, \mathbf{R}_2) - \circ\circ\circ,$$

$$n_1(\mathbf{R}_1)/f_0 = nf(\mathbf{R}_1) + \nu_1(\mathbf{R}_1) \quad (4.5)$$

$$- \int d^3R_2[\chi_2(\mathbf{R}_1, \mathbf{R}_2)n + \nu_1(\mathbf{R}_2)f(\mathbf{R}_1) + \nu_2(\mathbf{R}_1, \mathbf{R}_2)]$$

$$+ \int\int d^3R_2 d^3R_3[\chi_2(\mathbf{R}_1, \mathbf{R}_2)\nu_1(\mathbf{R}_3)$$

$$+ \tfrac{1}{2}f(\mathbf{R}_1)\nu_2(\mathbf{R}_2, \mathbf{R}_3) + \tfrac{1}{2}\chi_3(\mathbf{R}_1, \mathbf{R}_2, \mathbf{R}_3)n +$$

$$\tfrac{1}{2}\nu_3(\mathbf{R}_1, \mathbf{R}_2, \mathbf{R}_3)] - \circ\circ\circ.$$

These terms can be pictorially represented by diagrams as follows:

$$f_0 \quad = \quad \exp\left[\; - \; \blacktriangle \; + \; \tfrac{1}{2!} \; \blacktriangle\!\!-\!\!-\!\!\blacktriangle \; - \cdots \;\right]$$

$$n_0/f_0 \quad = \quad \circ \; - \; \circ\!\!-\!\!\cdot\!\!-\!\!\blacktriangle \; + \; \tfrac{1}{2} \; \cdots$$

$$n_1/f_0 \quad = \quad \circ \quad \triangle \quad + \quad \circ\!\!-\!\!\cdot\!\!-\!\!\triangle$$
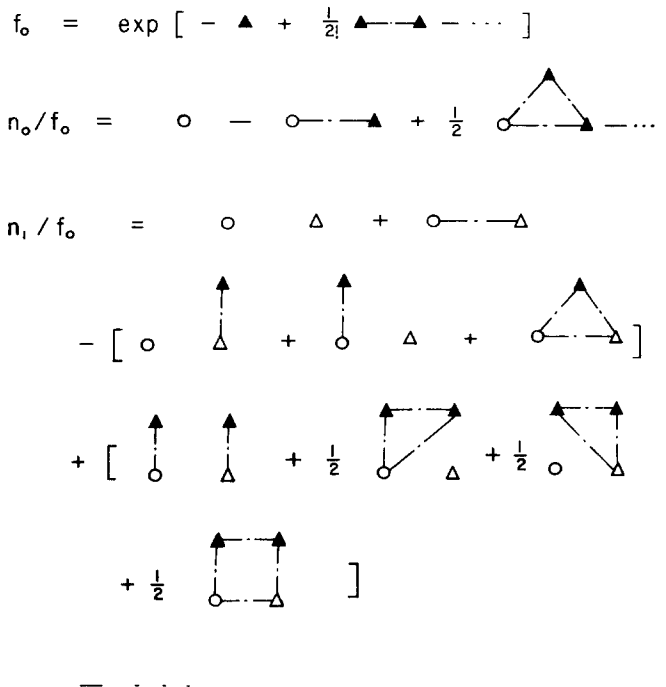
The rules of representation are:

(a) The electron is denoted by a circle and each impurity by a triangle.

(b) The dot—dash lines, called *correlation bonds*, denote correlation.

(c) If the particle variables correspond to the integration variables, the corresponding marks are blackened.

Important features of the diagrams for $\{n_{N_s}\}$ are

(a) Each diagram is a connected diagram whose components are all connected to the open marks corresponding to fixed particle coordinates.

(b) Diagrams are assigned with alternating signs as the number of impurities increases. In addition, a numerical factor is assigned which can be guessed from the diagram itself; if the interchange of blackened triangles brings back to itself, the symmetry number corresponding to all possible permutations is given.

A way of checking Eqs. (4.5) is to recalculate the reduced density operators $n^{(M)}$ through the original definitions (3.12) and (3.10). This exercise is however omitted here.

In normal applications of the theory, the impurity distribution is, by assumption, invariant under translation. This assumption in particular implies that the one-impurity distribution function $f(\mathbf{R})$ is constant, and is equal to n_s. While this assumption may simplify many formulas in the present section, we have proceeded in a most general manner; our theory can be applied to systems of arbitrary inhomogeneity and impurity correlation.

5. THE DIAGRAM REPRESENTATION

Let us consider N_s impurity ensemble. The Hamil-Hamiltonian

$$h_{N_s} \equiv h_0 + \lambda \sum_{\alpha=1}^{N_s} \tilde{v}_\alpha \qquad (5.1)$$

is independent of time. So is the corresponding Liouville operator ℓ_{N_s}. The formal solution of the Liouville equation (2.8) is

$$n_{N_s}(t) = \exp(-it\ell_{N_s})n_{N_s}(0) \equiv \exp(-it\ell_{N_s})n_{N_s}. \qquad (5.2)$$

In general a function of an operator is defined by a power series of a certain parameter, for example

$$\exp(-it\ell) = 1 - it\ell + \tfrac{1}{2}(-it)^2\ell^2 - \cdots. \qquad (5.3)$$

The same operator $\exp(-it\ell) = \exp[-it(\ell_0 + \lambda v)]$ can also be considered as a function of the coupling parameter λ, and can be expanded in a power series of λ as follows:

$$\exp(-it\ell) = \exp(-it\ell_0)$$
$$\times (1 + \sum_{1}^{\infty} (-i\lambda)^k \int_0^t d\tau_1 \int_0^{\tau_1} d\tau_2 \cdots$$
$$\int_0^{\tau_{k-1}} d\tau_k v(\tau_1)v(\tau_2)\cdots v(\tau_k)), \qquad (5.4)$$

where

$$v(\tau) \equiv \exp(i\tau\ell_0)v \exp(-i\tau\ell_0)$$
$$\equiv \exp(i\tau\ell_0)\sum_\alpha \tilde{v}_\alpha \exp(-i\tau\ell_0)$$
$$\equiv \sum_\alpha \tilde{v}_\alpha(\tau). \qquad (5.5)$$

We consider a grand ensemble which is characterized by a set of density operators $\{n_{N_s}\}$ at the initial time 0. Once these operators are given, they will evolve themselves in accordance with the Liouville equations (2.9), and each $n_{N_s}(t)$ will be given in the form (5.2). The set $\{n_{N_s}\}$ contains all the information about the initial condition of the system. As we saw in the last section, the same information can be given by the set of reduced density operators $\{n^{(M)}\}$ which are defined in (3.16). Alternatively, the same condition can be described by the set of correlation operators $\{n, \chi_l, \nu_k\}$. The descriptions by any of these three sets are equivalent. By choice we will specify the initial condition by the set of correlation operators. In terms of the latter, the N_s-impurity density operators are given in the form (4.4):

$$n_{N_s}(\mathbf{R}_1, \mathbf{R}_2, \ldots, \mathbf{R}_{N_s})$$
$$= n \prod_{1}^{N_s} f(\mathbf{R}_\alpha) + \{\nu_1(\mathbf{R}_1) \prod_{\alpha=2}^{N_s} f(\mathbf{R}_\alpha)$$

$$+ \text{ similar terms obtained by permutations}\} + \circ \circ \circ$$

$$- \{\text{terms involving the integration with respect to}$$

$$\mathbf{R}_{N_s+1}\} + \cdots. \qquad (5.6)$$

Let us now consider the reduced density operator averaged over impurity configurations

$$n(t) = \sum_{N_s=0}^{\infty} \frac{1}{N_s!} \int \circ \circ \circ \int d^3R_1 d^3R_2 \circ \circ \circ d^3R_{N_s} \exp(-it\ell_{N_s})n_{N_s}. \qquad (5.7)$$

We expand $\exp(-it\ell_{N_s})$ by means of (5.4) and n_{N_s} by (5.6). We then represent terms in the expansion by diagrams as follows:

We draw a horizontal solid line for the electron. Corresponding to $\tilde{v}_\alpha(\tau)$ [see Eq. (5.5)] we draw a dotted line line, called an *interaction bond*, connecting the impurity α and the electron at τ, $t > \tau > 0$, where the time is measured from the right to the left (Fig. 1). The ν_k are denoted by dot—dash lines connecting the right end of the electron line with k impurities. The χ_l are denoted by dot—dash lines connecting l impurities. In this way, we can represent all terms in the one-to-one correspondence.

Due to the fact that all ν_k and χ_l are symmetric, those diagrams which have the same structure but are distinguished only in having different impurity indices, contribute the same amount. We may thus stipulate that a diagram without indices represents the sum of the contributions of all impurity-indexed diagrams of the same structure.

6. DIAGRAM ANALYSIS FOR $n(t)$

A general diagram is composed of a number of sub-diagrams linked together by interaction and/or correlation bonds; one subdiagram contains the electron line.

Let us first consider the set of subdiagrams without the electron line. These diagrams are naturally time-independent, and are of the same structure as those appearing for f_0 in Eqs. (4.5). It can easily be verified that the totality of these subdiagrams is common to any linked subdiagram with the electron line, and that it contributes the factor

$$\exp[\int d^3R_1 f(\mathbf{R}_1) - \tfrac{1}{2}\int\int d^3R_1 d^3R_2 \chi_2(\mathbf{R}_1, \mathbf{R}_2) + \circ \circ \circ] = f_0^{-1}, \qquad (6.1)$$
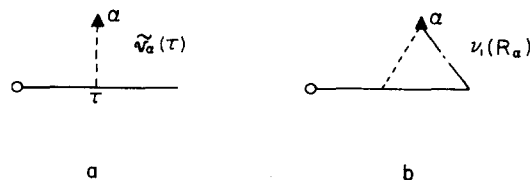
FIG. 1. Diagrams representing terms in the expansion of the time-dependent density operator $n(t)$. Diagram a contains the interaction bond representing $\tilde{\epsilon}_\alpha(\tau)$. Diagram b contains the correlation bond representing $\nu_1(\mathbf{R}_\alpha)$.

a   b
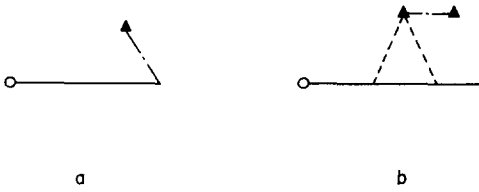
FIG. 2. Connected diagrams with correlation bonds ending at a blackened triangle do not contribute according to Theorem C.

which equals the inverse of f_0 in (4.5). This factor f_0^{-1} just compensates the factor f_0 which is common to all the expansions (4.5) of n_{N_s} $(N_s \geqslant 1)$ in terms of correlation operators. By this compensation, we may say that only the subdiagrams which each contain the electron line, called the *connected diagrams*, contribute to $n(t)$.

Let us now analyze these connected diagrams. Many diagrams contribute nothing because of the following Theorem C.

Theorem C: Any connected diagram which contains a correlation bond ending at a blackened triangle, contributes zero. Diagrams a and b in Fig. 2 belong to this class of noncontributing connected diagrams. The theorem may be verified by direct calculation. For example, diagram a appears twice in the series; it appears with the factor -1 from the expansion of n_0 and with the factor $+1$ from the expression n_1 [see Eq. (4.5)]; the total contribution thus vanishes. Such cancellation holds for all connected diagrams of this class.

After disregarding diagrams which do not contribute, we have yet to consider a large number of connected diagrams. A diagram will in general contain several free electron line segments. A line segment is said to be *free* if the diagram containing it is broken into two by cutting it. Free line segments are indicated by check marks $\sqrt{}$ in Fig. 3a. The diagram may contain a number of those parts which consist of nonfree line segments, interaction, and correlation bonds. Such a part will be called a *d-part* or *g*-part according to whether or not it contains a correlation bond. Diagram a in Fig. 3 contains a *d-part* and a *g*-part.

If a diagram should contain a *g*-part suspended by two free line segments or a *d*-part standing to the right of a free line segment, it could be *reduced* by suppressing the *g*- or *d*-part. In the process of reduction, the line segment marked by the open circle on the left should not be suppressed. With this rule, the reduction becomes unique. For example, in Fig. 3 diagram a can uniquely be reduced to diagram b. In fact, all connected diagrams (except one) are reducible to the single diagram b. Conversely, reducible diagrams can be obtained from diagram b by *dressing* its free electron line with *g*-and/or *d*-parts.

a   dress   b
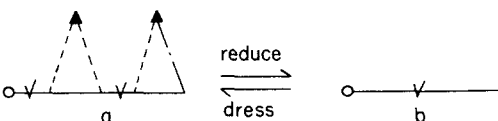
FIG. 3. Diagram a contains a *g*-part and a *d*-part, and it is reducible to diagram b; conversely it can be generated from diagram b by dressing with *g*- and *d*-parts.

The contribution of the irreducible diagram b is given by

$$\exp(-it\ell_0)n \equiv n_{\text{tree}}(t). \tag{6.2}$$

Clearly, this is the reduced density operator corresponding to the free-electron system. We have thus found that the single irreducible diagram b corresponds to $n_{\text{tree}}(t)$ while this diagram plus all reducible diagrams which may be obtained from it by dressing with *g*- and/or *d*-parts, correspond to $n(t)$.

7. DIAGRAM ANALYSIS FOR THE INTERACTION TERM; CLOSED EQUATION FOR $n(t)$

Let us consider the interaction term

$$-i\lambda \int d^3R[\bar{v}(\mathbf{r}-\mathbf{R}), n^{(1)}(\mathbf{R}, t)]$$

$$= (-i\lambda) \sum_{N_s=0}^{\infty} \mathrm{Tr}'_{,N_s}\{v_{N_s} \exp(-it\ell_{N_s})n_{N_s}\}, \tag{7.1}$$

which appeared in the hierarchy equation (3.24). After expanding both $\exp(-it\ell_{N_s})$ and n_{N_s}, we may represent terms in the expansion by diagrams. These diagrams will be different from those diagrams for $n(t)$ only in each having an interaction bond attached to the open circle, a bond representing v. Compare (7.1) with (5.7).

The diagrams can be analyzed in the same manner as before by using Theorem C, concepts of connected diagrams, reducibility, *g*- and *d*-parts, all of which hold in this case. Typical connected diagrams (which each contain open circles) are shown in Fig. 4. Diagram b, which contains two *g*-parts, is reducible to diagram a which has one *g*-part. It is easy to verify that any irreducible connected diagram has either *g*-part or *d*-part.

Let us first consider an irreducible diagram containing a *g*-part. Its contribution can always be expressed in the form of a certain function g of Liouville operators acting on the time-dependent reduced density operator corresponding to the free electron:

$$\mathrm{Tr}'\{g(\gamma)\}n_{\text{tree}}(t_\gamma), \tag{7.2}$$

where the symbol Tr' means the trace (integration) with respect to all impurities involved in the structure γ, and t_γ is the greatest time available to the free electron line. For example, the contribution of the irreducible diagram in Fig. 4(a) can be written down as

$$(-i\lambda)^2 \int d^3R_\alpha \, \bar{v}_\alpha \exp(-it\ell_0)\int_0^t d\tau \bar{v}_\alpha(\tau) \exp(i\tau\ell_0)n_{\text{tree}}(\tau). \tag{7.3}$$

Therefore, in this example

$$g(\gamma) = (-i\lambda)^2 \bar{v}_\alpha \exp(-it\ell_0) \int_0^t d\tau \bar{v}_\alpha(\tau) \exp(i\tau\ell_0),$$

$$t_\gamma = \tau. \tag{7.4}$$

a   b
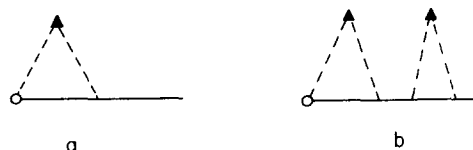
FIG. 4. Connected diagrams which contribute to the interaction term (7.1). Diagram b is reducible to diagram a whose contribution is given by (7.3).
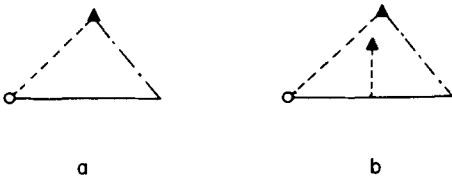
FIG. 5. Two irreducible connected diagrams with d-parts. The contribution of diagram a is given by (7.7).

Consider now reducible diagrams which upon reduction give rise to the irreducible diagram. The former can be constructed from the latter by dressing the free line on the right. The structures of all possible sub-diagrams which upon reduction give rise to the free line can be seen to be identical with those of the connected diagrams for $n(t)$, which we discussed in the last sections. These analyses lead us to write for the total contribution of the irreducible diagram containing a g-part and the reducible diagrams generated from it

$$\mathrm{Tr}'\{g(\gamma)\}n(t_\gamma),$$ (7.5)

which is obtained from (7.2) by simply suppressing the suffix "free." The total contribution of all the irreducible diagrams containing g-parts and reducible diagrams generated from them can now be obtained by summing over all possible irreducible structures γ:

$$\sum_\gamma \mathrm{Tr}'\{g(\gamma)\}n(t_\gamma).$$ (7.6)

Next, we consider irreducible diagrams containing d-parts. Two such diagrams are drawn in Fig. 5. Diagram a contributes

$$(-i\lambda) \int d^3R_\alpha \, \tilde{\upsilon}_\alpha \exp(-it\slashed{h}_0)\nu_1(\mathbf{R}_\alpha).$$ (7.7)

We note that the contribution can be expressed in the form of a function d of Liouville operators acting on the initial correlation operator ν. In fact, this feature is true for all diagrams containing single d-parts. We may therefore write the contribution of all the irreducible diagrams containing d-parts in the form

$$\sum \mathrm{Tr}'\{d\nu\}.$$ (7.8)

In summary, the interaction term (7.1) can be decomposed into the two sums (7.6) and (7.8). We can thus reexpress Eq. (3.24) in the following form:

$$[\partial/\partial t + i\slashed{h}_0]n(t) = \sum \mathrm{Tr}'\{g\}n + \sum \mathrm{Tr}'\{d\nu\}.$$ (7.9)

This is the equation desired. It is derived from the hierarchy equation (3.24) without introducing any approximation. Since the correlation operators ν are to be given as an initial condition, this equation contains only one unknown $n(t)$. It has, however, infinitely many terms; it is a linear and non-Markoffian equation.

8. REMARKS

Equation (7.9) has two sums $\sum \mathrm{Tr}'\{g\}n$ and $\sum \mathrm{Tr}'\{d\nu\}$, which will be referred to as collision and initial correlation terms, respectively. These terms can be written down by drawing, and reading off, irreducible connected diagrams. The diagrams are precisely those which defined the collision and initial correlation terms in the earlier finite-particle-number ensemble theory.[3,4] Thus, the present grand ensemble theory has

led to the same closed evolution equation without invoking the bulk limit, the use of which has clouded the earlier result as mentioned in Sec. 1.

Equation (7.9) has been studied by the present author and his coworkers. The main points in these studies will be summarized below. Let us first consider the collision terms. The operator $g(\gamma)$ depends on the impurity distribution. Let us suppose that this distribution is invariant under translation and without correlations:

$$f(\mathbf{R}_1) = \mathrm{const} = n_s,$$ (8.1)

$$\chi_l = 0, \quad l \geq 2.$$ (8.2)

This is the case for which almost all previous theories[2] were reported. In this case, the concentration dependence of $g(\gamma)$ is obtained by simply counting the number of impurity triangles in γ. [A similar consideration can be extended to initial correlation terms, too.] This is a remarkable property when we consider the fact that the the interaction term, Eq. (7.1), does not allow any simple concentration expansion. In other words, by the connected diagram analysis, we have not only achieved the rigorous transformation into a closed equation but also established the concentration expansion. Equation (7.9) can therefore be regarded as an ideal starting point from which one investigates the concentration expansion of a transport coefficient. In fact, using this equation, the formal density expansion of the electrical conductivity was reported earlier.[3,4]

It was soon found that this concentration expansion diverges term by term. This difficulty means that the summation of g- and d-diagrams according to the number of impurity triangles is not appropriate. The difficulty can be removed by resumming the series according to a different recipe, that is, in terms of proper connected diagrams.[6] The resummation was found to incorporate the so-called quasiparticle effect in a natural manner.

Equation (7.9) contains terms which explicitly depend on the initial condition. Many important physical properties of the system in or near equilibrium do not depend on the initial condition. Such examples are all equilibrium thermodynamic properties and transport coefficients. We may wonder whether there exist equations which are simpler than Eq. (7.9), which do not depend on the initial condition, but which nevertheless describe the equilibrium and near-equilibrium properties in a rigorous manner. The answer to this question is, in general, yes. In particular, it was shown earlier[7] that an equation for the asymptotic density operator $\bar{n}$ defined through

$$n(t) \to \bar{n} \quad \text{as} \quad t \to \infty$$ (8.3)

can be obtained from Eq. (7.9) (through the stationary-state approximation), that the obtained equation is independent of the initial condition, and that its solution yields the result identical with that of the correlation function formula for the electrical conductivity.[8]

The grand ensemble theory developed here can be adapted, with few changes, to treat a system of interacting particles which obey the Boltzmann statistics.[9] The quantum statistics, however, introduces fundamen-

tal changes in theory, and a simple extension of the theory to a quantum statistical gas or the electron-impurity system with the inclusion of the Coulomb interaction among electrons appears to be difficult.

ACKNOWLEDGMENTS

*Permanent address.

[1] e.g., A.H. Wilson, *Theory of Metals* (Cambridge U.P., Cambridge, 1936).

[2] L. Van Hove, Physica 21, 517 (1955); 23, 441 (1957); W. Kohn and J.M. Luttinger, Phys. Rev. 108, 590 (1957); J.M. Luttinger and W. Kohn, Phys. Rev. 109, 1892 (1958); S.F. Edwards, Phil. Mag. 3, 1020 (1958); G.V. Chester and A. Thellung, Proc. Phys. Soc. London 73, 765 (1959); 77, 1005 (1961); A. Janner, Helv. Phys. Acta 35, 45 (1962); I. Prigogine and P. Resibois, Physica 27, 29 (1961); A. Kawabata, J. Phys. Soc. Japan 23, 999 (1967); H. Scher and T. Holstein, Phys. Rev. 148, 598 (1966); A. Lodder and S. Fujita, J. Phys. Soc. Japan 25, 774 (1968).

[3] S. Fujita, *Introduction to Non-Equilibrium Quantum Statistical Mechanics* (Saunders, Philadelphia, 1966).

[4] S. Fujita, J. Phys. Chem. Solids 28, 615 (1966).

[5] J. Yvon, *Les Corrérations et l'entropie en mécanique statistique classique* (Dunod, Paris, 1966); Engl. transl.: J. Yvon, *Correlations and Entropy in Classical Statistical Mechanics*, translated by H.S.H. Massey (Pergamon, Oxford, 1969).

[6] S. Fujita and C.C. Chen, Intern. J. Theor. Phys. 2, 59 (1969).

[7] S. Fujita and A. Lodder, Physica 50, 541 (1970).

[8] R. Kubo, J. Phys. Soc. Japan 12, 570 (1957); H. Nakano, Prog. Theoret. Phys. 15, 77 (1956); M.S. Green, J. Chem. Phys. 22, 898 (1954).

[9] S. Fujita, Physica 76, 373, 395 (1974).

Sum rules for the optical constants

Frederick W. King*

Chemistry Department, Queen's University, Kingston, Ontario, Canada K7L 3N6
(Received 14 July 1975)

A number of sum rules for the optical constants, in particular, the refractive index of a nonconducting medium, are obtained. Some of the sum rule constraints are highly damped for large frequencies, exponentially in one particular case. Formal integral relationships for the index of refraction at complex frequencies are presented. Sum rules based on known experimental points for which $n(\omega)-1$ has zeros are indicated. An outline of a modified derivation of some recently presented sum rules for the optical constants is given.

I. INTRODUCTION

The Kramers-Kronig relations for linear optical properties have led to a number of useful sum rule constraints which these properties must satisfy. The underlying conditions for which these sum rules should hold are very general. The physical basis, the causality condition, determines the domain in which the generalized optical property is holomorphic, from which the Kramers-Kronig relations may then be deduced. A number of sum rules have been recently derived by employing the Kramers-Kronig relations and the appropriate asymptotic behavior of the particular optical properties at large frequencies. This asymptotic behavior may be readily determined by assuming at sufficiently high frequencies, the medium responds like a free electron gas.

This paper is concerned with the investigation of principally, the constraints on the generalized refractive index for a nonconducting medium. A number of surprisingly simple, but potentially very useful sum rules have been obtained for the generalized refractive index by Altarelli *et al.*[1,2] and Villani and Zimerman.[3] The simplest of these relations are

$$\int_0^\infty \omega \kappa(\omega)\, d\omega = \frac{\pi}{4}\, \omega_p^2, \tag{1}$$

$$\int_0^\infty [n(\omega) - 1]\, d\omega = 0, \tag{2}$$

$$\int_0^\infty [\omega \kappa(\omega)][n(\omega) - 1]\, d\omega = 0, \tag{3}$$

$$\cos\pi\beta \int_a^\infty \frac{[n(\omega) - 1]\, d\omega}{(\omega - a)^\beta(\omega + a)^\beta} - \sin\pi\beta \int_a^\infty \frac{\kappa(\omega)\, d\omega}{(\omega - a)^\beta(\omega + a)^\beta}$$

$$+ \int_0^a \frac{[n(\omega) - 1]\, d\omega}{(a - \omega)^\beta(a + \omega)^\beta} = 0, \quad -\tfrac{1}{2} < \beta < 1, \tag{4}$$

$$\int_0^\infty \omega^2[n(\omega) - 1]^2\, d\omega = \int_0^\infty \omega^2 \kappa^2(\omega)\, d\omega, \tag{5}$$

where $n(\omega)$ and $\kappa(\omega)$ are the real and imaginary parts, respectively, of the generalized refractive index, and ω_p is the plasma frequency. The first of these relationships, Eq. (1) is the well-known f sum rule, Eqs. (2) and (3) being derived by Altarelli *et al.*, and are designated as the ADNS sum rules, and the expressions given by Eqs. (4) and (5) have been obtained by Villani and Zimerman (designated VZ sum rules). Equation (4) contains the ADNS sum rule equation (2), as a special case; $\beta = 0$.

In Sec. II we outline how some of the above sum rules may be obtained in an alternative manner from the Kramers-Kronig relations. Section III presents a number of formal relationships connecting the generalized refractive index on the complex imaginary frequency axis, with the refractive index at real frequencies. In Sec. IV we outline some strikingly simple and apparently previously unnoticed sum rules based on the zeros of $n(\omega) - 1$. Section V deals with some generalizations of the above sum rules, which have the desirable property of being highly damped at large frequencies. Finally, in Sec. VI, we discuss possible extensions and limitations of the sum rules derived herein.

II. DERIVATION OF SUM RULES FROM THE KRAMERS-KRONIG RELATIONS

The basic equations from which most of the sum rules for the refractive index have been derived are the Kramers-Kronig relations

$$n(\omega_0) - 1 = \frac{2}{\pi} P \int_0^\infty \frac{\omega \kappa(\omega)\, d\omega}{\omega^2 - \omega_0^2}, \tag{6}$$

$$\kappa(\omega_0) = -\frac{2\omega_0}{\pi} P \int_0^\infty \frac{[n(\omega) - 1]\, d\omega}{\omega^2 - \omega_0^2}. \tag{7}$$

From Eq. (6), it follows immediately that

$$\int_0^\infty [n(\omega') - 1]\, d\omega' = \frac{2}{\pi} \int_0^\infty d\omega' \, P \int_0^\infty \frac{\omega \kappa(\omega)\, d\omega}{\omega^2 - \omega'^2} \tag{8}$$

and on interchanging the order of integration

$$\int_0^\infty [n(\omega') - 1]\, d\omega' = \frac{2}{\pi} \int_0^\infty \omega \kappa(\omega)\, d\omega \, P \int_0^\infty \frac{d\omega'}{\omega^2 - \omega'^2}, \tag{9}$$

and noting the relationship

$$P \int_0^\infty \frac{d\omega'}{\omega^2 - \omega'^2} = -\frac{\pi^2}{4}\, \delta(\omega), \tag{10}$$

then

$$\int_0^\infty [n(\omega') - 1]\, d\omega' = -\frac{\pi}{2} \int_0^\infty \omega \kappa(\omega)\delta(\omega)\, d\omega = 0, \tag{11}$$

which is the ADNS sum rule, Eq. (2). The conditions for the interchange of the order of integration to obtain Eq. (9) need to be scrutinized carefully. The necessary requirement is that the integrand, a function of ω and ω', be summable over the plane $-\infty < \omega < \infty$, $-\infty < \omega' < \infty$. Similarly, from the Kramers-Kronig relation (7),

$$\int_0^\infty \frac{\kappa(\omega')\,d\omega'}{\omega'} = -\frac{2}{\pi} \int_0^\infty d\omega' \, P \int_0^\infty \frac{[n(\omega) - 1]\,d\omega}{\omega^2 - \omega'^2}, \quad (12)$$

and by interchanging the order of integration, and employing Eq. (10), leads to the result

$$\int_0^\infty \frac{\kappa(\omega)\,d\omega}{\omega} = \frac{\pi}{2}[n(0) - 1]. \quad (13)$$

Depending on the behavior of $\kappa(\omega)$ as $\omega \to 0$, the integral in Eq. (13) may exist only in the sense of a principal values. Equation (13) is, of course, just the limit as $\omega_0 \to 0$ of the Kramers-Kronig relation, Eq. (6). It is worth pointing out that Eq. (13) is not applicable to metals where both $n(\omega)$ and $\kappa(\omega)$ behave as $\sim \omega^{-1/2}$ as $\omega \to 0$.

Similar derivations may be carried out by first squaring the appropriate Kramers-Kronig relations. From Eq. (7),

$$\frac{\kappa^2(\omega)}{\omega^2} = \frac{4}{\pi^2} P \int_0^\infty \frac{[n(\omega') - 1]\,d\omega'}{\omega'^2 - \omega^2}$$
$$\times P \int_0^\infty \frac{[n(\omega'') - 1]\,d\omega''}{\omega''^2 - \omega^2}, \quad (14)$$

and hence

$$\int_0^\infty \kappa^2(\omega)\,d\omega = \frac{4}{\pi^2} \int_0^\infty \omega^2\,d\omega \, P \int_0^\infty \frac{[n(\omega') - 1]\,d\omega'}{\omega'^2 - \omega^2}$$
$$\times P \int_0^\infty \frac{[n(\omega'') - 1]\,d\omega''}{\omega''^2 - \omega^2}. \quad (15)$$

Interchanging the order of integration as before, and making use of the result,

$$P \int_0^\infty \frac{\omega^2\,d\omega}{(\omega'^2 - \omega^2)(\omega''^2 - \omega^2)} = \frac{\pi^2}{4}[\delta(\omega' - \omega'') + \delta(\omega' + \omega'')] \quad (16)$$

converts Eq. (15) into

$$\int_0^\infty \kappa^2(\omega)\,d\omega = \int_0^\infty [n(\omega') - 1]\,d\omega'$$
$$\times \int_0^\infty [n(\omega'') - 1]\{\delta(\omega' - \omega'') + \delta(\omega' + \omega'')\}\,d\omega'', \quad (17)$$

and hence

$$\int_0^\infty \kappa^2(\omega)\,d\omega = \int_0^\infty [n(\omega) - 1]^2\,d\omega, \quad (18)$$

a result noted by Altarelli and Smith.[2] In a similar manner, squaring Eq. (6) and integrating over all frequencies gives

$$\int_0^\infty \omega^2[n(\omega) - 1]^2\,d\omega = \frac{4}{\pi^2} \int_0^\infty \omega^2\,d\omega \, P \int_0^\infty \frac{\omega'\kappa(\omega')\,d\omega'}{\omega'^2 - \omega^2}$$
$$\times P \int_0^\infty \frac{\omega''\kappa(\omega'')\,d\omega''}{\omega''^2 - \omega^2}. \quad (19)$$

Employing Eq. (16), we have the result

$$\int_0^\infty \omega^2[n(\omega) - 1]^2\,d\omega = \int_0^\infty \omega^2\kappa^2(\omega)\,d\omega,$$

which is the VZ sum rule, Eq. (5). The sum rule, Eq. (13), is obtained from the Kramers-Kronig relations as follows. Multiplying Eq. (6) by (7) gives the result

$$\frac{\kappa(\omega)[n(\omega) - 1]}{\omega} = -\frac{4}{\pi^2} P \int_0^\infty \frac{\omega'\kappa(\omega')\,d\omega'}{\omega'^2 - \omega^2}$$
$$\times P \int_0^\infty \frac{[n(\omega'') - 1]\,d\omega''}{\omega''^2 - \omega^2}, \quad (20)$$

and hence

$$\int_0^\infty \omega\kappa(\omega)[n(\omega) - 1]\,d\omega = -\frac{4}{\pi^2} \int_0^\infty \omega^2\,d\omega$$
$$\times P \int_0^\infty \frac{\omega'\kappa(\omega')\,d\omega'}{\omega'^2 - \omega^2} \cdot P \int_0^\infty \frac{[n(\omega'') - 1]\,d\omega''}{\omega''^2 - \omega^2}, \quad (21)$$

which yields

$$\int_0^\infty \omega\kappa(\omega)[n(\omega) - 1]\,d\omega = 0.$$

A number of additional results may be c' ined by considering higher powers of the generalized refractive index. The quantity $N^2(\omega) - 1$, where $N(\omega)$ is the generalized refractive index, is a holomorphic function in the upper half-plane, from which it can be deduced, with the appropriate asymptotic behavior, that

$$N^2(\omega') - 1 = \frac{P}{\pi i} \int_{-\infty}^\infty \frac{[N^2(\omega) - 1]\,d\omega}{\omega - \omega'}, \quad (22)$$

and hence

$$n^2(\omega') - \kappa^2(\omega') - 1 = \frac{4}{\pi} P \int_0^\infty \frac{\omega\kappa(\omega)n(\omega)\,d\omega}{\omega^2 - \omega'^2}, \quad (23)$$

$$n(\omega')\kappa(\omega') = -\frac{\omega'}{\pi} P \int_0^\infty \frac{[n^2(\omega) - \kappa^2(\omega) - 1]\,d\omega}{\omega^2 - \omega'^2}. \quad (24)$$

From Eqs. (9) and (23),

$$\int_0^\infty [n^2(\omega') - \kappa^2(\omega') - 1]\,d\omega' = \frac{4}{\pi} \int_0^\infty d\omega' \, P \int_0^\infty \frac{\omega\kappa(\omega)n(\omega)\,d\omega}{\omega^2 - \omega'^2}$$
$$= -\pi \int_0^\infty \omega\kappa(\omega)n(\omega)\delta(\omega)\,d\omega$$
$$= 0,$$

hence,

$$\int_0^\infty [n^2(\omega) - 1]\,d\omega = \int_0^\infty \kappa^2(\omega)\,d\omega. \quad (25)$$

Similarly, from Eqs. (9) and (24),

$$\int_0^\infty \frac{n(\omega)\kappa(\omega)\,d\omega}{\omega} = \frac{\pi}{4}[n^2(0) - 1]. \quad (26)$$

Equation (26) can also be recognized as the limit as $\omega' \to 0$ of the Kramers-Kronig relation, Eq. (23). Equation (26) is restricted to nonconductors. Sum rules for higher powers of the refractive index may also be obtained from the squares and the products of Eqs. (23) and (24). Thus from Eqs. (23) and (16), we get

$$\int_0^\infty [n^2(\omega) - \kappa^2(\omega) - 1]^2\omega^2\,d\omega = 4 \int_0^\infty \omega^2 n^2(\omega)\kappa^2(\omega)\,d\omega; \quad (27)$$

from Eqs. (24) and (16), we obtain

$$\int_0^\infty n^2(\omega)\kappa^2(\omega)\,d\omega = \frac{1}{4} \int_0^\infty [n^2(\omega) - \kappa^2(\omega) - 1]^2\,d\omega; \quad (28)$$

and from Eqs. (16), (23), and (24), we get

$$\int_0^\infty \omega n(\omega)\kappa(\omega)[n^2(\omega) - \kappa^2(\omega) - 1]\,d\omega = 0. \quad (29)$$

III. GENERALIZED REFRACTIVE INDEX AT COMPLEX FREQUENCIES

The generalized refractive index of a medium may be expressed in the form

$$N^2(\omega) = 1 + \int_0^\infty \exp(i\omega\tau)G(\tau)\,d\tau, \quad (30)$$

where the function $G(\tau)$ depends on the properties of the medium and the time, and ω is to be regarded as a complex frequency. From Eq. (30), it follows that for complex frequencies

$$N^2(-\omega^*) = N^{*2}(\omega). \tag{31}$$

For purely imaginary frequencies, $\omega = i\omega''$,

$$N^2(i\omega'') = N^{*2}(i\omega''), \tag{32}$$

and hence, on the purely imaginary frequency axis, $N^2(\omega)$ is real. A similar argument may be presented for the function $N(\omega) - 1$. The function $N^2(\omega) - 1$ obeys the relation[4]

$$\int_{-\infty}^{\infty} |N^2(\omega' + i\omega'') - 1|^2 \, d\omega < \text{const} \quad (\omega'' > 0), \tag{33}$$

which allows by use of Titchmarsh's theorem, for $N(\omega)$ to be written in the form

$$N(\omega) = 1 + \int_0^{\infty} \exp(i\omega\tau) F(\tau) \, d\tau, \tag{34}$$

where $F(\tau)$ depends on the properties of the medium; from which it follows that $N(\omega)$ is real on the imaginary frequency axis and in particular

$$\kappa(i\omega'') = 0 \tag{35}$$

for real ω''.

The simplest result for the refractive index at complex frequencies can be obtained by considering the integral of the function $[N(\omega) - 1]/(\omega - i\omega_0)$ around a half-circle contour containing the real axis, and the upper-half complex frequency plane. For ω_0 real and greater than zero, and the notation simplified by designating ω as a real variable, we have the result

$$\omega_0 \int_0^{\infty} \frac{[n(\omega) - 1] \, d\omega}{\omega^2 + \omega_0^2} + \int_0^{\infty} \frac{\omega\kappa(\omega) \, d\omega}{\omega^2 + \omega_0^2} = \pi[n(i\omega_0) - 1], \tag{36}$$

and considering the function $[N(\omega) - 1]/(\omega + i\omega_0)$, we obtain

$$\int_0^{\infty} \frac{\omega\kappa(\omega) \, d\omega}{\omega^2 + \omega_0^2} = \omega_0 \int_0^{\infty} \frac{[n(\omega) - 1] \, d\omega}{\omega^2 + \omega_0^2}. \tag{37}$$

From Eqs. (36) and (37), the results for the refractive index on the imaginary frequency axis are

$$\frac{\pi}{2}[n(i\omega_0) - 1] = \int_0^{\infty} \frac{\omega\kappa(\omega) \, d\omega}{\omega^2 + \omega_0^2}, \tag{38}$$

$$\frac{\pi}{2}[n(i\omega_0) - 1] = \omega_0 \int_0^{\infty} \frac{[n(\omega) - 1] \, d\omega}{\omega^2 + \omega_0^2}. \tag{39}$$

From Eqs. (38) and (39) the following results for integrals along the imaginary frequency axis can be obtained:

$$\int_0^{\infty} \frac{[n(i\omega') - 1] \, d\omega'}{\omega'} = \frac{2}{\pi} \int_0^{\infty} d\omega' \int_0^{\infty} \frac{[n(\omega) - 1] \, d\omega}{\omega^2 + \omega'^2}, \tag{40}$$

on interchanging the order of integration, Eq. (40) becomes

$$\int_0^{\infty} \frac{[n(i\omega') - 1] \, d\omega'}{\omega'} = \int_0^{\infty} \frac{[n(\omega) - 1] \, d\omega}{\omega}, \tag{41}$$

hence,

$$\int_0^{\infty} \frac{[n(i\omega) - n(\omega)] \, d\omega}{\omega} = 0; \tag{42}$$

and from Eq. (38),

$$\int_0^{\infty} [n(i\omega) - 1] \, d\omega = \int_0^{\infty} \kappa(\omega) \, d\omega. \tag{43}$$

A result analogous to Eq. (43), is well known for the dielectric constant.[5] Kramers-Kronig relations involving the refractive index on the imaginary axis may be derived by restricting the contour Γ to a circular arc in the first quadrant, then the integral

$$\int_{\Gamma} \frac{[N(\omega) - 1] \, d\omega}{\omega - \omega'}$$

(ω' on the real positive axis) leads to the results

$$\omega' \int_0^{\infty} \frac{[n(i\omega) - 1] \, d\omega}{\omega^2 + \omega'^2} = \pi[n(\omega') - 1] - P \int_0^{\infty} \frac{\kappa(\omega) \, d\omega}{\omega - \omega'}, \tag{44}$$

$$\int_0^{\infty} \frac{\omega[n(i\omega) - 1] \, d\omega}{\omega^2 + \omega'^2} = \pi\kappa(\omega') + P \int_0^{\infty} \frac{[n(\omega) - 1] \, d\omega}{\omega - \omega'}. \tag{45}$$

Sum rules which are more rapidly convergent for large frequencies can be generated by noting values of the derivatives of the generalized refractive index on the imaginary frequency axis. The integral

$$\oint \frac{[N(\omega) - 1] \, d\omega}{(\omega^2 + 1)^2}$$

evaluated for a semicircular contour in the upper-half complex plane yields the result

$$\int_0^{\infty} \frac{[n(\omega) - 1] \, d\omega}{(\omega^2 + 1)^2} = \frac{\pi}{4}[n(i) - 1] - \frac{\pi i}{4} \left(\frac{d[N(\omega) - 1]}{d\omega} \right)_{\omega=i}; \tag{46}$$

and since

$$\left(\frac{d[N(\omega) - 1]}{d\omega} \right)_{\omega=i} = \frac{4i}{\pi} \int_0^{\infty} \frac{\omega\kappa(\omega) \, d\omega}{(\omega^2 + 1)^2}, \tag{47}$$

Eq. (46) simplifies to

$$\int_0^{\infty} \frac{(1 - \omega^2)[n(\omega) - 1] \, d\omega}{(\omega^2 + 1)^2} = 2 \int_0^{\infty} \frac{\omega\kappa(\omega) \, d\omega}{(\omega^2 + 1)^2}. \tag{48}$$

Equation (48) may be readily checked by differentiating the integrals in Eq. (37) with respect to ω_0, which leads to

$$\int_0^{\infty} \frac{\omega\kappa(\omega) \, d\omega}{(\omega^2 + \omega_0^2)^2} = \frac{1}{2\omega_0} \int_0^{\infty} \frac{[n(\omega) - 1](\omega_0^2 - \omega^2) \, d\omega}{(\omega^2 + \omega_0^2)^2}, \tag{49}$$

and thus reduces to Eq. (48) on setting $\omega_0 = 1$. Considering in a similar manner, the integral

$$\oint \frac{[N(\omega) - 1] \, d\omega}{(\omega^2 + 1)^3},$$

leads to the result

$$\int_0^{\infty} \frac{[n(\omega) - 1] \, d\omega}{(\omega^2 + 1)^3} = \frac{3\pi}{16}[n(i) - 1]$$

$$- \frac{3\pi i}{16} \left(\frac{d[N(\omega) - 1]}{d\omega} \right)_{\omega=i} - \frac{\pi}{16} \left(\frac{d^2[N(\omega) - 1]}{d\omega^2} \right)_{\omega=i}. \tag{50}$$

With the result

$$\left(\frac{d^2[N(\omega) - 1]}{d\omega^2} \right)_{\omega=i} = \frac{4}{\pi} \int_0^{\infty} \frac{[n(\omega) - 1](3\omega^2 - 1) \, d\omega}{(\omega^2 + 1)^3}, \tag{51}$$

Eq. (50) simplifies to the same result as was obtained by considering the integral

$$\oint \frac{[N(\omega) - 1]\, d\omega}{(\omega + 1)^2},$$

i.e.,

$$\int_0^\infty \frac{[n(\omega) - 1](1 - \omega^2)\, d\omega}{(\omega^2 + 1)^2} = 2\int_0^\infty \frac{\omega\kappa(\omega)\, d\omega}{(\omega^2 + 1)^2}. \qquad (52)$$

Equation (52) may be rearranged with the aid of Eq. (2) to the alternative form

$$2\int_0^\infty \frac{\omega\kappa(\omega)\, d\omega}{(\omega^2 + 1)^2} = -\int_0^\infty \frac{\omega^2(\omega^2 + 3)[n(\omega) - 1]\, d\omega}{(\omega^2 + 1)^2}. \qquad (53)$$

IV. ZEROS OF THE FUNCTION [n(ω) - 1]

From the ADNS sum rule, Eq. (2), it is obvious that there must exist at least one zero of the function $[n(\omega) - 1]$ [except in the trivial and unrealistic situation that $n(\omega) = 1$ for all ω]. If this zero is designated by the point $\omega = \omega_0$, then from the Kramers-Kronig relationship, Eq. (6), we have the immediate result that

$$\int_0^\infty \frac{\omega\kappa(\omega)\, d\omega}{\omega^2 - \omega_0^2} = 0. \qquad (54)$$

The necessity of retaining the principal value in Eq. (54) depends on the behavior of $\kappa(\omega)$ at the point $\omega = \omega_0$.

From the Kramers-Kronig relations, subtracted dispersion relations for the real and imaginary parts of the generalized refractive index can be written

$$n(\omega') - n(\omega'') = \frac{2}{\pi}(\omega'^2 - \omega''^2)P\int_0^\infty \frac{\omega\kappa(\omega)\, d\omega}{(\omega^2 - \omega'^2)(\omega^2 - \omega''^2)}, \qquad (55)$$

$$\kappa(\omega') - \kappa(\omega'') = \frac{2}{\pi}(\omega'' - \omega')P\int_0^\infty \frac{[n(\omega) - 1][\omega^2 + \omega'\omega'']\, d\omega}{(\omega^2 - \omega'^2)(\omega^2 - \omega''^2)}. \qquad (56)$$

For any two given frequencies ω_0', ω_0'' $(\omega_0' \neq \omega_0'')$, which are both zeros of the function $[n(\omega) - 1]$, or alternatively for which $[n(\omega_0') - n(\omega_0'')]$ vanishes, then from Eq. (55),

$$P\int_0^\infty \frac{\omega\kappa(\omega)\, d\omega}{(\omega^2 - \omega_0'^2)(\omega^2 - \omega_0''^2)} = 0. \qquad (57)$$

Similarly, if there exist two distinct frequencies ω_1, ω_2 $(\omega_1 \neq 0, \ \omega_2 \neq 0)$ such that $[\kappa(\omega_1) - \kappa(\omega_2)] = 0$, then Eq. (56) yields the result

$$P\int_0^\infty \frac{[n(\omega) - 1][\omega^2 + \omega_1\omega_2]\, d\omega}{(\omega^2 - \omega_1^2)(\omega^2 - \omega_2^2)} = 0. \qquad (58)$$

The relationship Eq. (57) also follows directly from Eq. (54) when two zeros of the function $[n(\omega) - 1]$ are known. In fact, for a set of distinct frequencies which are zeros of $[n(\omega) - 1]$, a compact form for the sum rules can be written as

$$\int_0^\infty \omega\kappa(\omega)\, d\omega = \int_0^\infty \frac{\omega^3\kappa(\omega)\, d\omega}{\omega^2 - \omega_0^2} = \int_0^\infty \frac{\omega^3\kappa(\omega)\, d\omega}{\omega^2 - \omega_1^2} = \cdots, \qquad (59)$$

where $\omega_0, \omega_1, \cdots$ designate the zeros of $[n(\omega) - 1]$. If we employ the f sum rule [Eq. (1)], Eq. (59) becomes

$$\frac{\pi}{4}\omega_p^2 = \int_0^\infty \frac{\omega^3\kappa(\omega)\, d\omega}{\omega^2 - \omega_0^2} = \int_0^\infty \frac{\omega^3\kappa(\omega)\, d\omega}{\omega^2 - \omega_1^2} = \cdots. \qquad (60)$$

V. HIGHLY DAMPED CONSTRAINTS

The desirability of finding highly damped sum rule relations for the optical constants lies in the importance of deemphasizing certain frequency regions. For the high-frequency region, the optical constants are not ascertainable with high accuracy. The purpose of this section is to outline some possibly useful constraints for providing a consistency check of experimental values, while at the same time, damping the high-frequency results, so that inaccuracies in this region may be ignored. The results obtained make use of some of the formal relations for the refractive index at imaginary frequency.

The method of generating highly damped sum rules consists of considering the appropriate analytic exponential function $[\exp(a\omega)/(\exp(b\omega) + 1)]\exp(i\delta\omega)$ $(\delta > 0; \ b > a; \ a, \ b, \ \delta \text{ all real})$ multiplied by the function $[N(\omega) - 1]$. Considering the integral

$$\oint \frac{\exp(a\omega)}{\exp(b\omega) + 1}\exp(i\delta\omega)[N(\omega) - 1]\, d\omega,$$

with the contour a semicircle in the upper half-plane,

$$\oint \frac{\exp(a\omega)}{\exp(b\omega) + 1}\exp(i\delta\omega)[N(\omega) - 1]\, d\omega$$

$$= 2\pi i\Sigma\left(\text{residues at } \frac{n\pi i}{b}, n = 1, 3, \cdots\right), \qquad (61)$$

then using Jordan's Lemma, and taking the limit $\delta \to + 0$,

$$\int_{-\infty}^\infty \frac{\exp(a\omega)}{\exp(b\omega) + 1}[N(\omega) - 1]\, d\omega$$

$$= -\frac{2\pi i}{b}\sum_{j=0}^\infty \exp\left[(ia\pi/b)(2j + 1)\right]\left[N\left((2j + 1)\frac{\pi i}{b}\right) - 1\right]. \qquad (62)$$

Separating Eq. (61) into real and imaginary parts, leads to the results

$$\int_{-\infty}^\infty \frac{\exp(a\omega)}{\exp(b\omega) + 1}[n(\omega) - 1]\, d\omega$$

$$= \frac{2\pi}{b}\sum_{j=0}^\infty \sin\left(\frac{\pi a}{b}(2j + 1)\right)\left[n\left((2j + 1)\frac{\pi i}{b}\right) - 1\right] \qquad (63)$$

and

$$\int_{-\infty}^\infty \frac{\exp(a\omega)\kappa(\omega)\, d\omega}{\exp(b\omega) + 1}$$

$$= -\frac{2\pi}{b}\sum_{j=0}^\infty \cos\left(\frac{\pi a}{b}(2j + 1)\right)\left[n\left((2j + 1)\frac{\pi i}{b}\right) - 1\right]. \qquad (64)$$

Some special cases of Eqs. (63) and (64) can be given. Taking the $\lim a \to + 0$ for Eq. (63), leads to the result

$$\int_0^\infty [n(\omega) - 1]\, d\omega = 0,$$

which is the ADNS sum rule. Taking the $\lim a \to + 0$ for Eq. (64) and making use of Eq. (38) leads to a trivial identity [both sides equal to $-\int_0^\infty \kappa(\omega)\tanh(\frac{1}{2}b\omega)\, d\omega$]. This identity can be proved easily for general a, b ($b > a$) by employing Eq. (38) and the result

$$\sum_{j=0}^\infty \frac{\cos[(\pi a/b)(2j + 1)]}{(2j + 1)^2 + b^2\omega^2\pi^{-2}}$$

$$= \frac{\pi^2}{4b\omega}[\cosh a\omega \tanh \tfrac{1}{2}b\omega - \sinh a\omega]. \qquad (65)$$

The $\lim b \to a$ leads to similar results. An interesting special case is obtained for $2a = b$. From Eqs. (63) and (38) we have

$$\int_0^\infty [n(\omega) - 1] \operatorname{sech}\left(\frac{b\omega}{2}\right) d\omega$$

$$= \frac{4}{b} \sum_{j=0}^\infty (-1)^j \int_0^\infty \frac{\omega \kappa(\omega)\, d\omega}{\omega^2 + (2j+1)^2 \pi^2 b^{-2}}$$

$$= \left(\frac{4b}{\pi^2}\right) \int_0^\infty \omega \kappa(\omega)\, d\omega \sum_{j=0}^\infty \frac{(-1)^j}{(2j+1)^2 + b^2 \omega^2 \pi^{-2}} . \quad (66)$$

Now employing the result

$$\sum_{j=0}^\infty \frac{(-1)^j}{(2j+1)^2} = \beta(2) \quad \text{[Catalan's constant (Ref. 6)]}, \quad (67)$$

where $\beta(2) \approx 0.91596$, and the inequaltiy

$$\frac{4b}{\pi^2} \beta(2) \int_0^\infty \omega \kappa(\omega)\, d\omega > \frac{4b}{\pi^2} \int_0^\infty \omega \kappa(\omega)$$

$$\times \left(\sum_{j=0}^\infty \frac{(-1)^j}{(2j+1)^2 + b^2\omega^2\pi^{-2}}\right) d\omega, \quad 0 < b < \infty \quad (68)$$

allows the following inequality to be written:

$$b\beta(2)\pi^{-1}\omega_p^2 > \int_0^\infty [n(\omega) - 1]\operatorname{sech}\left(\frac{b\omega}{2}\right) d\omega, \quad (69)$$

which is a principal result of this paper. The integral in Eq. (69) is very strongly damped at high frequencies. The special case $2a = b$ for Eq. (64) leads to a trivial identity.

A general extension of the above procedure consists of considering the situation in which the appropriate function has a singularity on the real axis, in addition to the singularities on the imaginary frequency axis. Consider the integral

$$\oint \frac{\exp(a\omega)\exp(i\delta\omega)}{\exp(b\omega)+1} \frac{[N(\omega)-1]\, d\omega}{\omega - \omega_0}$$

($\delta > 0$; $b > a$: a, b, δ, and ω_0 real), with the contour taken as a semicircle in the upper half-plane, including the real axis, with the contour indented (into the upper complex plane) at the singularity on the real axis. The limit $\delta \to +0$ gives

$$P\int_{-\infty}^\infty \frac{\exp(a\omega)[N(\omega)-1]\, d\omega}{[\exp(b\omega)+1](\omega-\omega_0)} = -2\pi i \sum_{j=0}^\infty \frac{\exp[(a/b)\pi i(2j+1)]}{\pi i(2j+1) - b\omega_0}\left[N\left(\frac{\pi i}{b}(2j+1)\right)-1\right] + \frac{i\pi \exp(a\omega_0)}{\exp(b\omega_0)+1}[N(\omega_0)-1]. \quad (70)$$

The real and imaginary parts are

$$P\int_{-\infty}^\infty \frac{[n(\omega)-1]\exp(a\omega)\, d\omega}{[\exp(b\omega)+1](\omega-\omega_0)} + \frac{\pi\exp(a\omega_0)\kappa(\omega_0)}{[\exp(b\omega_0)+1]} = -2\sum_{j=0}^\infty \frac{\{n[(\pi i/b)(2j+1)]-1\}}{(2j+1)^2 + b^2\omega_0^2\pi^{-2}}$$

$$\times \left[(2j+1)\cos\left(\frac{\pi a}{b}(2j+1)\right) + b\omega_0\pi^{-1}\sin\left(\frac{\pi a}{b}(2j+1)\right)\right] \quad (71)$$

and

$$P\int_{-\infty}^\infty \frac{\kappa(\omega)\exp(a\omega)\, d\omega}{[\exp(b\omega)+1](\omega-\omega_0)} - \frac{\pi[n(\omega_0)-1]\exp(a\omega_0)}{[\exp(b\omega_0)+1]} = -2\sum_{j=0}^\infty \frac{\{n[(\pi i/b)(2j+1)]-1\}}{(2j+1)^2 + b^2\omega_0^2\pi^{-2}}$$

$$\times \left[(2j+1)\sin\left(\frac{\pi a}{b}(2j+1)\right) - b\omega_0\pi^{-1}\cos\left(\frac{\pi a}{b}(2j+1)\right)\right]. \quad (72)$$

Some interesting special cases can be obtained from Eqs. (71) and (72). Taking the $\lim a \to +0$ in Eqs. (71) and (72) leads to the results

$$P\int_0^\infty \frac{[n(\omega)-1]\omega\tanh(\tfrac{1}{2}b\omega)\, d\omega}{\omega^2 - \omega_0^2} + \frac{\pi}{2}\kappa(\omega_0)\tanh\left(\frac{b\omega_0}{2}\right) = 2\sum_{j=0}^\infty \frac{(2j+1)\{n[(\pi i/b)(2j+1)]-1\}}{(2j+1)^2 + b^2\omega_0^2\pi^{-2}}, \quad (73)$$

$$P\int_0^\infty \omega_0 \frac{\kappa(\omega)\tanh(\tfrac{1}{2}b\omega)\, d\omega}{\omega^2 - \omega_0^2} - \frac{\pi}{2}\tanh\left(\frac{b\omega_0}{2}\right)[n(\omega_0)-1] = -2b\omega_0\pi^{-1}\sum_{j=0}^\infty \frac{n[(\pi i/b)(2j+1)]-1}{(2j+1)^2 + b^2\omega_0^2\pi^{-2}}. \quad (74)$$

Combining Eqs. (73) and (38) and employing the result

$$\tfrac{7}{8}\zeta(3) = \sum_{j=0}^\infty \frac{1}{(2j+1)^3} > \sum_{j=0}^\infty \frac{2j+1}{(2j+1)^2 + b^2\omega_0^2\pi^{-2}} \frac{1}{(2j+1)^2 + b^2\omega^2\pi^{-2}}, \quad (75)$$

where $\zeta(m)$ is Riemann's zeta function $[\zeta(3) \approx 1.202]$, leads to the result

$$P\int_0^\infty \frac{\omega\tanh(\tfrac{1}{2}b\omega)[n(\omega)-1]\, d\omega}{\omega^2 - \omega_0^2} < \tfrac{7}{8}\zeta(3)b^2\omega_p^2\pi^{-2} - \tfrac{1}{2}\pi\kappa(\omega_0)\tanh\left(\frac{b\omega_0}{2}\right). \quad (76)$$

Combining Eqs. (73) and (39) leads directly to the inequality

$$\tfrac{1}{2}\pi\kappa(\omega_0)\tanh\left(\frac{b\omega_0}{2}\right) + P\int_0^\infty \frac{\omega\tanh(\tfrac{1}{2}b\omega)[n(\omega)-1]\, d\omega}{\omega^2 - \omega_0^2} < 0, \quad (77)$$

where the result

$$\sum_{j=0}^\infty \frac{1}{(2j+1)^2} > \sum_{j=0}^\infty \frac{(2j+1)^2}{(2j+1)^2 + b^2\omega_0^2\pi^{-2}} \frac{1}{(2j+1)^2 + b^2\omega^2\pi^{-2}} \quad (78)$$

and the ADNS sum rule, Eq. (2), have been employed. Equation (77) is a stronger inequality than Eq. (76) for the upper bound to the integral. Similar results can be derived from Eq. (74):

$$P \int_0^\infty \frac{\tanh(b\omega/2)\kappa(\omega)\,d\omega}{\omega^2 - \omega_0^2} + \tfrac{15}{16} b^3 \pi^{-3} \omega_p^2 \, \zeta(4) > \tfrac{1}{2}\pi \omega_0^{-1} \tanh\left(\frac{b\omega_0}{2}\right) [n(\omega_0) - 1] \tag{79}$$

and

$$\omega_0 P \int_0^\infty \frac{\tanh(b\omega/2)\kappa(\omega)\,d\omega}{\omega^2 - \omega_0^2} > \tfrac{1}{2}\pi \tanh\left(\frac{b\omega_0}{2}\right) [n(\omega_0) - 1], \tag{80}$$

where $\zeta(4) \approx 1.082$. Equation (80) is the stronger inequality for a lower bound to the integral

$$P \int_0^\infty \frac{\tanh(b\omega/2)\kappa(\omega)\,d\omega}{\omega^2 - \omega_0^2}.$$

VI. DISCUSSION

The simplified procedure for derivation of many of the sum rules for the optical constants discussed in Sec. II is restricted only by the necessity of establishing the conditions of summability of the integrands, so that, the order of integration may be inverted. Summability encompasses the asymptotic behavior that has been assumed in the derivation of these sum rules by Altarelli et al.[1]

The establishment of results for the optical constants as a function of imaginary frequency are of little utility from the point of view of providing a constraint for the testing of experimental data. However, they do provide a route to other sum rules, which evolve from particular functional forms of the optical constants for which poles on the imaginary axis appear, as in the case of the integrals considered in Sec. V.

Constraints on the optical constants which arise from the zeros of the particular optical functions have possible wide utility. The majority of the known sum rule constraints, with the principal exceptions of Eqs. (1) and (2), relate the integrals of different optical constants over an infinite frequency interval. In order to test experimental data by such sum rules, both optical constants need to be experimentally accessible, or obtainable through indirect means (i.e., by Kramers–Kronig inversions). Sum rules of the type (54) circumvent this difficulty, since only experimental results in the vicinity of the zero(s) need to be determined for one of the optical constants, and not the entire frequency interval of the particular optical constant. This represents a considerable reduction in effort compared with determining data over large frequency intervals.

The key results of Sec. V, Eqs. (69), (77), and (80) provide useful results from the point of view of providing suitable criteria for the quality of optical data. The highly damped nature of Eq. (69) essentially eliminates the difficulty of obtaining data at high frequencies, or

having only poor data available for this region. Attempts to obtain strongly damped integral constraints as a function of a single optical constant have not as yet met with success.

The generalization of the above sum rules for conductors can be carried out in a straightforward manner. Connections between the various optical constants, such as the dielectric constant, the conductivity, etc., also allow a number of sum rules to be readily obtained. Extensions of the asymptotic method and the above procedures to second-order Kramers–Kronig relations for nonlinear optical phenomena[7-9] appear to be possible and this will be the subject of a further investigation.

ACKNOWLEDGMENTS

*Present address: Physical Chemistry Laboratory, Oxford University, South Parks Rd., Oxford OX1 3QZ, U.K.
[1]M. Altarelli, D. L. Dexter, H. M. Nussenzveig, and D. Y. Smith, Phys. Rev. B 6, 4502 (1972). An earlier brief application taking advantage of the limiting behavior of the frequency dependence of the optical constants as $\omega \to \infty$ is, W. M. Saslow, Phys. Lett. A 33, 157 (1970).
[2]M. Altarelli and D. Y. Smith, Phys. Rev. B 9, 1290 (1974).
[3]A. Villani and A. H. Zimerman, Phys. Rev. B 8, 3914 (1973).
[4]H. M. Nussenzveig, Causality and Dispersion Relations (Academic, New York, 1972), Chap. 1.
[5]L. D. Landau and E. M. Lifshitz, Electrodynamics of Continuous Media (Pergamon, New York, 1960), p. 262. See also, Statistical Physics (Pergamon, London, 1959), Sec. 122.
[6]Handbook of Mathematical Functions, edited by M. Abramowitz and I. A. Stegun (Dover, New York, 1970), p. 807.
[7]F. Smet and P. Smet, Nuovo Cimento B 20, 273 (1974).
[8]M. Asdente, M. C. Pascucci, A. M. Ricca, and P. Venturini, Solid State Commun. 15, 1551 (1974).
[9]F. L. Ridener, Jr. and R. H. Good, Jr., Phys. Rev. B 10, 4980 (1974).

Ruelle's cluster property for nonstrictly localizable fields derived from symmetry of their Wightman functions

W. Lücke

Institut für Theoretische Physik der Technischen Universität Clausthal, Clausthal-Zellerfeld, Germany
(Received 14 August 1975)

Hermitian fields are considered which fulfill all the Wightman axioms, except local commutativity, formulated for the Gel'fand space $S^1(\mathbb{R}^4)$ instead of the Schwartz space $\mathcal{S}(\mathbb{R}^4)$. For fields of this class, containing the Jaffe class and even the larger class of essentially local fields on $S^1(\mathbb{R}^4)$, it is shown that symmetry of the Wightman functions and mass gap imply Ruelle's cluster property. Hence the Haag–Ruelle–Hepp scattering formalism applies to these fields, which need not be local.

1. INTRODUCTION

While there is no doubt that relativistic quantum fields have to be treated as operator valued *generalized* functions,[1] there are no physical arguments for the choice of any particular space of test functions. It is mainly a matter of convenience[2] that the Schwartz spaces $\mathcal{S}$ and $\mathcal{D}$ have become standard in axiomatic quantum field theory.[3] In order to be as general as possible, it would be desirable to work with test spaces as small as possible and endowed with a topology as fine as possible. However, there naturally occur difficulties if the space of test functions is chosen too small. Perhaps the best known effect of this type is the following one analyzed by Jaffe[4]:

If one is to allow the field operators to increase rapidly at infinity in momentum space, it is natural to choose some space of smooth momentum space test functions $\tilde{\varphi}(p)$ which sufficiently rapidly decrease at infinity. However, such a function cannot have a Fourier transform $\varphi(x)$ with compact support (in configuration space) if it decreases too rapidly, i.e., if

$$\int_1^\infty dt\, t^{-2} \log(1/\max_{\|q\|>t} |\tilde{\varphi}(q)|)$$

diverges. Then the usual formulation of the axiom of local commutativity[3] (in short: *locality*), which is the most powerful tool in axiomatic quantum field theory, is no more applicable.

In order to allow for both locality and rapid increase of the field operators in momentum space, Jaffe introduced a new class of test spaces.[4] For fields of the Jaffe class most of the results of Wightman field theory can still be derived by essentially the same methods. Unfortunately, there is no Jaffe space minimal in the sense specified above. As a consequence, each dynamics requires its own Jaffe space. In this connection the Gel'fand space[5] $S^1(\mathbb{R}^4) \equiv S^{1,1,1,1}$ becomes relevant, as already realized by Constantinescu.[6] S^1 is just the intersection of all Jaffe spaces,[7] and its topology is finer than that induced by any Jaffe space. Consequently, it seems plausible that quite a lot of the structure of Jaffe spaces survives for S^1.

Therefore, in the present paper, we study fields fulfilling all the Wightman axioms[3] formulated for $S^1(\mathbb{R}^4)$ instead of $\mathcal{S}(\mathbb{R}^4)$, with the unavoidable exception of locality. Evidently, this class is considerably larger

than the Jaffe class.[8] For simplicity we consider only a Hermitian field $A(x)$ describing a single type of neutral scalar particle with mass $m > 0$. Constantinescu remarked[6] that it might be possible to derive many of the standard results of local quantum field theory also for such a field if "local commutativity is replaced with some 'technical' conditions."

We believe that symmetry of the Wightman functions, formulated in Sec. 2 as "Wightman property" and well known for local fields, is the natural and powerful substitute for locality. Obviously, PCT symmetry can then be proved by exactly the same methods as those used for Wightman fields.[3] However, it is by no means evident how to consistently define an S-matrix under these conditions. A solution of this problem seems to be of considerable interest in connection with the Osterwalder–Schrader approach[9] to constructive quantum field theory.

Now, the purpose of the present paper is to derive Ruelle's cluster property[10] from the usual mass gap condition and the Wightman property. Then we already know how to establish the Haag–Ruelle–Hepp scattering formalism.[11]

2. THE WIGHTMAN FUNCTIONS AND THEIR PROPERTIES

Let D be the common invariant dense domain of the smeared field operators[3]

$$A(\varphi) \equiv \int dx\, A(x)\varphi(x), \quad \varphi \in S^1(\mathbb{R}^4).$$

Since for S^1 there is an analog[12] of the classical nuclear theorem,[13] the expectation values

$$\langle \Phi | A(x_1) \cdots A(x_n) | \Psi \rangle, \quad \Phi, \Psi \in D,$$

and especially the vacuum expectation values

$$\mathfrak{W}(\hat{x}) \equiv \langle \Omega | A(x_1) \cdots A(x_n) | \Omega \rangle$$

$[\hat{x} \equiv (x_1, \ldots, x_n)]$ are well defined as generalized functions on $S^1(\mathbb{R}^{4n}) \equiv S^{1,\cdots,1}$. The same holds for the truncated vacuum expectation values, formally defined by

$$\mathfrak{W}^T(\hat{x}) \equiv \sum_{l=1} (-1)^{l-1}(l-1)! \sum_{M \in P_l(n)} \prod_{J \in M} \mathfrak{W}(x_{j_1}, \ldots, x_{j_{k_J}}),$$

where $M = \{J_1, \ldots, J_l\} \in P_l(n)$ iff it is a partition of $\{1, \ldots, n\}$ into l (nonempty) disjoint ordered subsets J_r with the ordering in each $J = (j_1, \ldots, j_{k_J})$ being the

natural relative ordering of integers. By translation in-variance of the theory, there are[14] generalized func-tions $W^{(T)}(\hat{\xi})$ on $S^1(\mathbb{R}^{4n})$ such that[15]

$$\int dx_1 \cdots dx_{n+1} \mathfrak{W}^{(T)}(x_1, \ldots, x_{n+1}) \varphi(x_1, \ldots, x_{n+1})$$

$$= \int d\hat{\xi}\, W^{(T)}(\xi) \left[\int d\xi_{n+1}\, \varphi\left(\sum_{j=1}^{n+1} \xi_j, \ldots, \sum_{j=n+1}^{n+1} \xi_j \right) \right]$$

$(d\hat{\xi} \equiv d\xi_1 \cdots d\xi_n,\ \xi_j \equiv x_j - x_{j+1},\ x_{n+2} \equiv 0)$ holds for every $\varphi \in S^1(\mathbb{R}^{4(n+1)})$, which we symbolically denote by

$$W^{(T)}(x_1 - x_2, \ldots, x_n - x_{n+1}) = \mathfrak{W}^{(T)}(x_1, \ldots, x_{n+1}).$$

In the usual way[3] one derives

$$\mathrm{supp}\,\widetilde{W^{(T)}}(\hat{q}) \subset (\overline{V_+})^n$$

(V_+ denotes the forward lightcone, $\overline{V_+}$ its closure) for their Fourier transforms, formally defined by

$$\widetilde{W^{(T)}}(\hat{q}) \equiv (2\pi)^{-2n} \int d\hat{\xi}\, W^{(T)}(\hat{\xi})\, \exp(i\hat{q}\hat{\xi})$$

$(\hat{q}\hat{\xi} = q_1\xi_1 + \cdots + q_n\xi_n,\ q_j\xi_j \equiv q_j^0 \xi_j^0 - \mathbf{q}_j \boldsymbol{\xi}_j)$. Therefore,[16] the Wightman functions

$$\underline{W}^{(T)}(\hat{\xi}) \equiv (2\pi)^{-2n} \int d\hat{q}\, \widetilde{W^{(T)}}(\hat{q})\, \exp(-i\hat{q}\hat{\xi})$$

are well-defined analytic functions, holomorph in the tube

$$T_n \equiv \{\hat{\xi} = \hat{\xi} - i\hat{\eta} \in \mathbb{C}^{4n} \colon \eta_j \in V_+ \text{ for } j = 1, \ldots, n\}.$$

$L_+^!$-invariance of the Wightman distributions implies $L_+^!$-invariance of their Fourier transforms and finally of the Wightman functions:

$$\underline{W}^{(T)}(\hat{\xi}) = \underline{W}^{(T)}(\Lambda\hat{\xi}) \tag{2.1}$$

for all $\Lambda \in L_+^!$. Hence the BHW theorem[3] tells us that the $\underline{W}^{(T)}(\hat{\xi})$ possess single valued analytic continuations into the "extended tube"

$$T_n' \equiv \bigcup_{\Lambda \in L_+(\mathbb{C})} \Lambda T_n,$$

where (2.1) also holds for $\Lambda \in L_+(\mathbb{C})$. The point $\hat{\xi}$ is known to be in $T_n' \cap \mathbb{R}^{4n}$, the set of so-called *Jost points*, iff[3] $\lambda_1\xi_1 + \cdots + \lambda_n\xi_n$ is spacelike for arbitrary nonnega-tive $\lambda_1, \ldots, \lambda_n$ with $\lambda_1 + \cdots + \lambda_n > 0$.

The Wightman functions are polynomially bounded[16] if $\hat{\eta}$ is restricted to a compact subset K of $(V_+)^n$:

$$\left| \underline{W}^{(T)}(\hat{\xi}) \right| < P_K(\hat{\xi}) \text{ for all } \hat{\xi} \in T_n \text{ with } \hat{\eta} \in K. \tag{2.2}$$

(Here P_K denotes a suitable polynomial depending on K.) Hence[17] the application

$$\varphi \to \lim_{\substack{\|\hat{\eta}\| \to 0 \\ \hat{\eta} \in (V_+)^n}} \int d\hat{\xi}\, \underline{W}^{(T)}(\hat{\xi}) \varphi(\hat{\xi}), \quad \varphi \in S^1(\mathbb{R}^{4n})$$

defines a generalized function on $S^1(\mathbb{R}^{4n})$, which coin-cides with the corresponding Wightman distribution:

$$\lim_{\substack{\|\hat{\eta}\| \to 0 \\ \hat{\eta} \in (V_+)^n}} \int d\hat{\xi}\, \underline{W}^{(T)}(\hat{\xi}) \varphi(\hat{\xi}) = \int d\hat{\xi}\, W^{(T)}(\hat{\xi}) \varphi(\hat{\xi}).$$

For the vacuum expectation values $(n > 1)$ this means

$$\int d\hat{x}\, \mathfrak{W}^{(T)}(\hat{x}) \varphi(\hat{x}) = \lim_{\substack{|\mathrm{Im}(z_k - z_{k+1})| \to 0 \\ \hat{z} \in \mathfrak{S}_n}} \int d\hat{x}$$

$$\times \underline{\mathfrak{W}}^{(T)}(\hat{z}) \varphi(\hat{x}) \text{ for } \varphi \in S^1(\mathbb{R}^{4n}), \tag{2.3}$$

where

$$\mathfrak{S}_n \equiv \{\hat{z} \in \mathbb{C}^{4n} \colon (z_1 - z_2, \ldots, z_{n-1} - z_n) \in T_{n-1}\},$$

$$\underline{\mathfrak{W}}^{(T)}(\hat{z}) \equiv \underline{W}^{(T)}(z_1 - z_2, \ldots, z_{n-1} - z_n).$$

This is why analysis of the Wightman functions gives deep insight into the general structure of the underlying field theory. The most useful tool in this connection is the following property of the Wightman functions, which we shall call the "Wightman property":

Definition: The Wightman functions are said to possess the *Wightman property* iff, for arbitrary $n > 1$, $\underline{\mathfrak{W}}^{(T)}(\hat{z})$ has a single valued analytic continuation into the domain[23]

$$\mathfrak{S}_n^P \equiv \{\hat{z} \in \mathbb{C}^{4n} \colon (z_{\pi(1)} - z_{\pi(2)}, \ldots, z_{\pi(n-1)} - z_{\pi(n)})$$

$$\in T_{n-1}' \text{ for suitable } \pi \in S_n\},$$

in which it is invariant under permutations of the $z_1, \ldots, z_n$.

The Wightman property is known to be equivalent to locality for tempered fields.[18] The same holds for Jaffe fields.[4] An obvious advantage of the Wightman property is that it does not depend on the test space on which $A(x)$ can be defined. While the usual formulation of locality fails for the Gel'fand space S^1, since it does not contain test functions with compact support, the formulation of the Wightman property still applies. Thus, in a technically modified way, the notation of locality becomes applicable to fields on $S^1(\mathbb{R}^4)$.

3. THE WIGHTMAN PROPERTY DERIVED FROM ESSENTIAL LOCALITY

In a recent paper[11] another substitute for locality, called "essential locality," was proposed. In this con-nection it is interesting to note that the Wightman prop-erty can also be derived from essential locality. This derivation will be given in the present section as an exercise for our main task, namely, the proof of Ruelle's cluster property in the next section. Let us first briefly review the concept of essential locality[19]:

A set $S \subset S^1(\mathbb{R}^n)$ is bounded with respect to the topo-logy of $S^1(\mathbb{R}^n)$ iff there is a positive constant A such that[20]

$$\sup_{\varphi \in S} \sup_{\chi \in M} \sup_{\alpha \in \mathbb{Z}_+^n} A^{-|\alpha|} \alpha^{-\alpha} \|\chi\|^N \left| \varphi^{(\alpha)}(\chi) \right|$$

$$< \infty \text{ for every } N \in \mathbb{Z}_+ \tag{3.1}$$

holds with $M = \mathbb{R}^n$. If (3.1) holds for some closed subset M of $\mathbb{R}^n$, then S will be called *locally bounded* on M in $S^1(\mathbb{R}^n)$. This definition was introduced in Ref. 11 in a slightly different form. Actually, both definitions are equivalent, due to the following

Lemma 3.1: Let O be some nonempty open subset of $\mathbb{R}^n$, let N be some nonnegative integer, and let ϵ, A be positive constants fulfilling the inequality $\epsilon < (2eA)^{-1}$. Finally, let $\varphi(\chi)$ be an arbitrarily differentiable function over O, for which

$$C_N \equiv \sup_{\chi \in O} \sup_{\alpha \in \mathbb{Z}_+^n} A^{-|\alpha|} \alpha^{-\alpha} (1 + \|\chi\|)^N \left| \varphi^{(\alpha)}(\chi) \right|$$

is finite. Then $\varphi(\chi)$ is the restriction of a function $\varphi(\sigma)$ that is holomorph in $U_\epsilon^c(O)$ and fulfils the inequality[21]

$$\sup_{\sigma \in U_\epsilon^c(0)} \sup_{\alpha \in Z_+^n} (2A)^{-|\alpha|} \alpha^{-\alpha} (1 + \|\sigma\|)^N |\varphi^{(\alpha)}(\sigma)|$$

$$\leq (1+\epsilon)^N C_N e^n (1 - 2eA\epsilon)^{-n}.$$

Proof: Evidently, the analytic continuation of $\varphi(\chi)$ is given by

$$\varphi^{(\alpha)}(\chi + \sigma') = \sum_{\beta \in Z_+^n} (\beta!)^{-1} \sigma'^\beta \varphi^{(\alpha+\beta)}(\chi)$$

for $\chi \in O$ and $\sigma' \in \mathbb{C}^n$ with $\|\sigma'\| < \epsilon$.

By this representation we get the estimate

$$(1 + \|\chi + \sigma'\|)^N |\varphi^{(\alpha)}(\chi + \sigma')|$$
$$\leq (1 + \|\sigma'\|)^N \sum_{\beta \in Z_+^n} (\beta!)^{-1} \|\sigma'\|^{|\beta|} (1 + \|\chi\|)^N |\varphi^{(\alpha+\beta)}(\chi)|$$
$$\leq (1 + \|\sigma'\|)^N C_N \sum_{\beta \in Z_+^n} (\beta!)^{-1} \|\sigma'\|^{|\beta|} A^{|\alpha+\beta|} (\alpha+\beta)^{(\alpha+\beta)}.$$

Hence the simple inequalities

$$(\alpha+\beta)^{(\alpha+\beta)} \leq 2^{|\alpha+\beta|} \alpha^\alpha \beta^\beta, \tag{3.2}$$
$$\beta! \geq e^{-n} e^{-|\beta|} \beta^\beta$$

yield

$$(1 + \|\chi + \sigma'\|)^N |\varphi^{(\alpha)}(\chi + \sigma')|$$
$$\leq (1 + \|\sigma'\|)^N C_N e^n (2A)^{|\alpha|} \alpha^\alpha \sum_{\beta \in Z_+^n} (2eA\|\sigma'\|)^{|\beta|}$$

for $\chi \in O$ and $\sigma' \in \mathbb{C}^n$ with $\|\sigma'\| < \epsilon$. This directly implies the statement of Lemma 3.1. ∎

Let M be a closed subset of $\mathbb{R}^n$ and $F(\chi)$ a generalized function on $S^1(\mathbb{R}^n)$. Then F is called *locally continuous* on M with respect to $S^1(\mathbb{R}^n)$ iff[11] $\sup_{\varphi \in S} |F(\varphi)|$ is finite for every $S \subset S^1(\mathbb{R}^n)$ that is locally bounded on M in $S^1(\mathbb{R}^n)$.

The field $A(x)$ is called *essentially local* iff,[11] for arbitrary $\Phi, \Psi \in D$, $\langle \Phi | [A(x_1), A(x_2)] | \Psi \rangle$ is locally continuous on $V_8 \equiv \{(x_1, x_2) \in \mathbb{R}^8 : (x_1 - x_2)^2 \geq 0\}$. In order to derive the Wightman property from essential locality, we first need one more technical result:

Lemma 3.2: Let C be a nonempty, open, convex cone in $\mathbb{R}^n$. Let the function $f(\chi - i\tau)$ be holomorph in $\mathbb{R}^n - iC$ and polynomially bounded on every region of the form $\mathbb{R}^n - iK$, where K is any compact subset of C. Finally, let $\varphi(\chi)$ be a test function from $S^1(\mathbb{R}^n)$ fulfilling

$$\sup_{\chi \in \mathbb{R}^n} \sup_{\alpha \in Z_+^n} A^{-|\alpha|} \alpha^{-\alpha} (1 + \|\chi\|)^N |\varphi^{(\alpha)}(\chi)| < \infty$$

for all $N \in Z_+$ and some $A > 0$ not depending on N. Then for arbitrary $\tau' \in C$ with $\|\tau'\| < (2eA)^{-1}$ we have

$$\lim_{\substack{\|\tau\| \to 0 \\ \tau \in C}} \int d\chi f(\chi - i\tau) \varphi(\chi) = \int d\chi f(\chi - i\tau') \varphi(\chi - i\tau').$$

Proof: Let $\tau' \in C$ and $\|\tau'\| < (2eA)^{-1}$. Then for sufficiently small fixed $\tau \in C$ we may choose $\epsilon > 0$ and a finite set $\{\tau_0, \ldots, \tau_N\} \subset C$ with $\tau_0 = \tau$, $\tau_N = \tau'$ such that the conditions

$$\|\tau_{j+1} - \tau_j\| < \epsilon, \quad U_\epsilon(\tau_j) \subset C, \quad \|\tau_j - \tau\| + \epsilon < (2eA)^{-1}$$

hold for $j = 0, 1, \ldots, N-1$. Under these conditions

Lemma 3.1 and the Cauchy theorem imply

$$\int d\chi f(\chi - i\tau_j) \varphi(\chi - i(\tau_j - \tau))$$
$$= \int d\chi f(\chi - i\tau_{j+1}) \varphi(\chi - i(\tau_{j+1} - \tau))$$

for $j = 0, \ldots, N-1$. Hence we conclude

$$\int d\chi f(\chi - i\tau) \varphi(\chi) = \int d\chi f(\chi - i\tau') \varphi(\chi - i(\tau' - \tau))$$

for sufficiently small $\tau \in C$. Obviously, by Lemma 3.1, the rhs converges to

$$\int d\chi f(\chi - i\tau') \varphi(\chi - i\tau')$$

for $\tau \to 0$. ∎

Let us define sequences of "type δ" by[22]

$$\varphi_{\hat{x}', N}(\hat{x}) \equiv (N/\pi)^{n/2} \exp(-N\|\hat{x} - \hat{x}'\|^2),$$

where $\hat{x}, \hat{x}' \in \mathbb{R}^{4n}$ and $N - 1 \in Z_+$. By Eqs. (2.2) and (2.3) and by Lemma 3.2 we then conclude:

$$\lim_{N \to \infty} \Big(\int d\hat{x} \underline{\mathfrak{W}}^{(T)}(\hat{x}) \varphi_{\hat{x}', N}(\hat{x})$$
$$- \int_{K_{\hat{x}'}^\epsilon} d\hat{x} \underline{\mathfrak{W}}^{(T)}(\hat{x} - i\hat{y}) \varphi_{\hat{x}', N}(\hat{x} - i\hat{y}) = 0$$

for sufficiently small $\hat{y} \in \mathbb{R}^{4n}$ with $(y_1 - y_2, \ldots, y_{n-1} - y_n) \in (V_+)^{n-1}$, where

$$K_{\hat{x}'}^\epsilon \equiv \{\hat{x} \in \mathbb{R}^{4n} : |x_j^r - x_j'^r| < \epsilon$$
$$\text{for } j = 1, \ldots, n \text{ and } r = 0, \ldots, 3\}.$$

If $\hat{x} \in \mathfrak{S}_n' \cap \mathbb{R}^{4n}$ and if we choose ϵ sufficiently small, then by the Cauchy theorem we may derive

$$\lim_{N \to \infty} \int_{K_{\hat{x}'}^\epsilon} d\hat{x} \underline{\mathfrak{W}}^{(T)}(\hat{x} - i\hat{y}) \varphi_{\hat{x}', N}(\hat{x} - i\hat{y}) = \underline{\mathfrak{W}}^{(T)}(\hat{x}')$$

for sufficiently small $\hat{y} \in \mathbb{R}^{4n}$ with $(y_1 - y_2, \ldots, y_{n-1} - y_n) \in (V_+)^{n-1}$. Hence

$$\lim_{N \to \infty} \int d\hat{x} \underline{\mathfrak{W}}^{(T)}(\hat{x}) \varphi_{\hat{x}', N}(\hat{x}) = \underline{\mathfrak{W}}^{(T)}(\hat{x}')$$

holds for every $\hat{x} \in \mathfrak{S}_n' \cap \mathbb{R}^{4n}$. If[23] $\pi \in S_n$ is such that also $\hat{x}'_\pi \equiv (x'_{\pi(1)}, \ldots, x'_{\pi(n)}) \in \mathfrak{S}_n'$, then, using

$$\varphi_{\hat{x}', N}(\hat{x}) = \varphi_{\hat{x}'_\pi, N}(\hat{x}_\pi),$$

we may conclude that

$$\lim_{N \to \infty} \int d\hat{x} [\underline{\mathfrak{W}}^{(T)}(\hat{x}) - \underline{\mathfrak{W}}^{(T)}(\hat{x}_\pi)] \varphi_{\hat{x}', N}(\hat{x})$$
$$= \underline{\mathfrak{W}}^{(T)}(\hat{x}') - \underline{\mathfrak{W}}^{(T)}(\hat{x}'_\pi). \tag{3.3}$$

Recalling the classification of Jost points we see that

$$\hat{x}' \in \mathbb{R}^{4n} \backslash V_{4n} \equiv \{\hat{x} \in \mathbb{R}^{4n} : (x_j - x_{j'})^2 < 0 \text{ for } j \neq j'\}.$$

Therefore, the set

$$\{N \varphi_{\hat{x}', N}(\hat{x})\}_{N \in Z_+}$$

is locally bounded on V_{4n} with respect to $S^1(\mathbb{R}^{4n})$. This may be easily proved by (3.2) and the estimate

$$A^{-(j+k+1)} k^{-k/2} j^{-j/2} \left| t^k \left(\frac{d}{dt}\right)^j \exp(-t^2) \right| < 1$$

for $t \in \mathbb{R}^1$ and $j, k \in Z_+$,

which holds[24] for sufficient large $A > 0$ and implies

$$N A^{-(j+k+1)} \epsilon^{j+2} [(j+k+2)^{-(j+k+2)/2} j^{-j/2}$$
$$\times \left| t^k \left(\frac{d}{dt}\right)^j \exp(-Nt^2) \right| < 1$$

for $N, j, k \in \mathbf{Z}_+$ and $t \in \mathbb{R}^1$ with $|t| > \epsilon$.

If $A(x)$ is essentially local, then

$$\sup_{\varphi \in S} \left| \int d\hat{x} \, [\underline{\mathfrak{W}}^{(T)}(\hat{x}) - \underline{\mathfrak{W}}^{(T)}(\hat{x}_\tau)] \varphi(\hat{x}) \right|$$

is finite for every set $S \subset S^1(\mathbb{R}^{4n})$ that is locally bounded on V_{4n} in $S^1(\mathbb{R}^{4n})$. Hence (3.3) finally yields

$$\underline{\mathfrak{W}}^{(T)}(\hat{x}') = \underline{\mathfrak{W}}^{(T)}(\hat{x}'_\tau)$$

for every $\hat{x}' \in \mathfrak{S}'_n \cap \mathbb{R}^{4n}$ and every $\pi \in S_n$ with $\hat{x}'_\tau \in \mathfrak{S}'_n$. This is known[25] to imply the Wightman property. Thus we have shown:

If $A(x)$ is essentially local, then its Wightman functions possess the Wightman property!

4. RUELLE'S CLUSTER PROPERTY AS A CONSEQUENCE OF WIGHTMAN PROPERTY AND MASS GAP

We do not know, whether, conversely, essential locality can be derived from the Wightman property. However, we shall derive some weaker continuity property, which, together with the mass gap condition, is still strong enough to guarantee Ruelle's cluster property.[10] To this end we have to extend the bound (2.2) to larger regions. This may be done using $L_+(\mathbb{C})$-invariance:

Lemma 4.1: Let $\lambda_1 \in (0,1)$, let $\lambda_2, \epsilon > 0$, and let $\eta'_1, \ldots, \eta'_{n-1} \in V_+$. Then, for sufficiently small $\delta > 0$, there is a compact subset K of $(V_+)^n$ and a finite constant C such that, with $\eta'_n \equiv 0$, the following statement holds:

For every $\hat{\eta} \in U_\delta(\hat{\eta}')$ and for every $\hat{\xi} \in \mathbb{R}^{4n}$ with

$$|\xi^0_n| < \lambda_1 \|\xi_n\| > \epsilon, \quad \|\xi_n\| > \lambda_2 \|\xi_r\| \quad \text{for } r = 1, \ldots, n-1$$

there is a $\Lambda \in L_+(\mathbb{C})$ fulfilling

$$\Lambda \hat{\xi} \in \mathbb{R}^{4n} - iK$$

and

$$\|\operatorname{Re}(\Lambda \hat{\xi})\| < \|\hat{\xi}\| + C$$

(remember $\zeta_r \equiv \xi_r - i\eta_r$).

Proof: Evidently, without loss of generality we may impose the additional restrictions

$$\xi_n = (a, b, 0, 0), \quad |a| < \lambda_1 b > \epsilon.$$

Let us first choose a $\delta_1 > 0$ such that

$$\overline{U_{3\delta_1}(\eta'_r)} \subset V_+ \quad \text{for } r = 1, \ldots, n-1.$$

Then choose a $\kappa > 0$ small enough in order to guarantee $\kappa / \lambda_2 < \delta_1$, $\kappa \lambda_1 / \epsilon < 1$ and $\operatorname{Re}(\Lambda_{\kappa, \epsilon_n} \eta'_r) \in U_{\delta_1}(\eta'_r)$ for $r = 1, \ldots, n-1$ and all ξ_n under consideration, where

$$\Lambda_{\kappa, \epsilon_n} \equiv \begin{pmatrix} (1 - \kappa^2 \|\xi_n\|^{-2})^{1/2} & -i\kappa \|\xi_n\|^{-1} & 0 & 0 \\ -i\kappa \|\xi_n\|^{-1} & (1 - \kappa^2 \|\xi_n\|^{-2})^{1/2} & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}$$

$$\in L_+(\mathbb{C}).$$

Finally, choose $\delta \in (0, \delta_1)$ sufficiently small to be sure

that the compact set

$$K \equiv \{\hat{\eta}'' \in \mathbb{R}^{4n} : \eta''_r \in \overline{U_{3\delta_1}(\eta'_r)} \text{ for } r = 1, \ldots, n-1;$$

$$\eta''_n \in \overline{U_\delta((b'', a'', 0, 0))} \text{ for suitable } a'', b''$$

$$\text{fulfilling } |a''| < \lambda_1 b'' \text{ and } \|(b'', a'', 0, 0)\| = \kappa\}$$

is contained in $(V_+)^n$. Then it is almost trivial to check

$$\Lambda_{\kappa, \epsilon_n} \hat{\xi} \in \mathbb{R}^{4n} - iK$$

and

$$\|\operatorname{Re}(\Lambda_{\kappa, \epsilon_n} \hat{\xi})\| < \|\hat{\xi}\| + \delta + \sum_{r=1}^{n-1} \|\eta'_r\|$$

for all $\hat{\xi}$ under consideration (corresponding to our choice for δ). ∎

Given ϵ, λ_1, $\lambda_2 > 0$ with $\lambda_1 < 1$ and given n, $j \in Z_+$ with $0 < j < n$, let us define

$$O^{n,j}_{\epsilon, \lambda_1, \lambda_2} \equiv \{x \in \mathbb{R}^{4n} : |x^0_j - x^0_{j+1}| < \lambda_1 \|\mathbf{x}_j - \mathbf{x}_{j+1}\| > \epsilon,$$

$$\|\mathbf{x}_j - \mathbf{x}_{j+1}\| > \lambda_2 \|x_r - x_{r'}\| \text{ if } \{r, r'\} \neq \{j, j+1\}\},$$

$$M^{n,j}_{\epsilon, \lambda_1, \lambda_2} \equiv \mathbb{R}^{4n} \backslash O^{n,j}_{\epsilon, \lambda_1, \lambda_2}.$$

Clearly, $\hat{x} \in O^{n,j}_{\epsilon, \lambda_1, \lambda_2}$ iff $\hat{x}_{\pi_j} \in O^{n,j}_{\epsilon, \lambda_1, \lambda_2}$, where π_j is the permutation which just exchanges j and $j+1$. In view of the cluster property we want to prove that

$$\sup_{\varphi \in S} \left| \int d\hat{x} [\mathfrak{W}^T(\hat{x}) - \mathfrak{W}^T(\hat{x}_{\tau j})] \varphi(\hat{x}) \right| < \infty \tag{4.1}$$

holds for every set $S \subset S^1(\mathbb{R}^{4n})$ fulfilling

$$\sup_{\varphi \in S} \sup_{\hat{z} \in U_\epsilon(M^{n,j}_{\epsilon, \lambda_1, \lambda_2})} \sup_{\alpha \in \mathbf{z}^{4n}_+} \epsilon^{|\hat{\alpha}|} \hat{\alpha}^{-\hat{\alpha}} \|\hat{z}\|^N |\varphi^{(\hat{\alpha})}(\hat{z})|$$

$$< \infty \text{ for } N \in \mathbf{Z}_+ \tag{4.2}$$

and

$$\sup_{\hat{z} \in U_\epsilon(\mathbb{R}^{4n})} \sup_{\alpha \in \mathbf{z}^{4n}_+} (2e\epsilon)^{|\hat{\alpha}|} \hat{\alpha}^{-\hat{\alpha}} \|\hat{z}\|^N |\varphi^{(\hat{\alpha})}(\hat{z})| \tag{4.3}$$

$$< \infty \text{ for every } N \in \mathbf{Z}_+ \text{ and } \varphi \in S.$$

Consider a set S of this type and let $\varphi \in S$. Then, by Lemma 3.2 in connection with (2.2) and (2.3), we have

$$\int d\hat{x} [\mathfrak{W}^T(\hat{x}) - \mathfrak{W}^T(\hat{x}_{\tau j})] \varphi(\hat{x})$$

$$= \int d\hat{x} \, \mathfrak{W}^T(\hat{z}) [\varphi(\hat{z}) - \varphi(\hat{z}_{\tau j})], \quad \hat{z} \equiv \hat{x} - i\hat{y}, \tag{4.4}$$

for $\hat{y} \in \mathbb{R}^{4n}$ with $\|\hat{y}\| < \epsilon$ and $y_r - y_{r+1} \in V_+$ for $r = 1, \ldots, n-1$. In order to apply the Wightman property, we should like to change the way of integration such that the $\hat{x} \in \mathbb{R}^{4n} \backslash U_\epsilon(M^{n,j}_{\epsilon, \lambda_1, \lambda_2})$ appear only in the combination $\hat{x} - i\hat{y}'$, where $\hat{y}' \in \mathbb{R}^{4n}$ is chosen (independently of φ) in such a way that

$$y'_j = y'_{j+1}, \quad \|\hat{y}'\| < \epsilon, \quad \text{and} \quad y'_r - y'_{r+1} \in V_+ \text{ for } r \neq j.$$

By Lemma 4.1, (2.2), and $L_+(\mathbb{C})$-invariance of W^T we know that there is a $\delta > 0$ and a polynomial $P(\hat{x})$ such that $\hat{z} \in \mathfrak{S}'_n$ and $\mathfrak{W}^T(\hat{z})$ fulfills the inequality

$$|\mathfrak{W}^T(\hat{z})| < P(\hat{x}) \tag{4.5}$$

for all $\hat{z}$ with $\hat{y} \in U_\delta(\hat{y}')$ and $\hat{x} \in U_\delta(O^{n,j}_{\epsilon, \lambda_1, \lambda_2})$. Therefore, let us put

$$y_r = y'_r \text{ for } r \neq j$$

and

$$y_j = (y_j'^0 + \min\{\epsilon - \|\hat{y}'\|, \delta\}/2, y_j')$$

in (4.4), which we are certainly allowed to do, if δ is chosen sufficiently small. Then, from the bound (2.2) and from (4.2), we see that

$$\left| \int_{M_{\epsilon,\lambda_1,\lambda_2}^{n,j}} d\hat{x}\, \underline{\mathfrak{W}}^T(\hat{z})[\varphi(\hat{z}) - \varphi(\hat{z}_{\tau_j})] \right|$$

has an upper bound independent of $\varphi \in S$. Hence, in order to establish (4.1), it is sufficient to show that also

$$\left| \int_{O_{\epsilon,\lambda_1,\lambda_2}^{n,j}} dx\, \underline{\mathfrak{W}}^T(\hat{z})[\varphi(\hat{z}) - \varphi(\hat{z}_{\tau_j})] \right|$$

has an upper bound independent of $\varphi \in S$. Let us define

$$\hat{O}_{\epsilon,\lambda_1,\lambda_2}^{n,j} \equiv \{(x_1, \ldots, \mathbf{x}_j, \ldots, x_n) \in \mathbb{R}^{4n-1}:$$

$$\hat{x} \in O_{\epsilon,\lambda_1,\lambda_2}^{n,j} \text{ for suitable } x_j^0 \in \mathbb{R}^1\}.$$

For $(x_1, \ldots, \mathbf{x}_j, \ldots, x_n) \in \hat{O}_{\epsilon,\lambda_1,\lambda_2}^{n,j}$ and $x_j^0 \in \mathbb{R}^1$ we then have:

$$\hat{x} \in O_{\epsilon,\lambda_1,\lambda_2}^{n,j} \text{ iff } x_{j+1}^0 - \lambda_1 \|\mathbf{x}_j - \mathbf{x}_{j+1}\|$$

$$< x_j^0 < x_{j+1}^0 + \lambda_1 \|\mathbf{x}_j - \mathbf{x}_{j+1}\|.$$

Therefore, using the Cauchy theorem, we may write

$$\int_{O_{\epsilon,\lambda_1,\lambda_2}^{n,j}} d\hat{x}\, \underline{\mathfrak{W}}^T(\hat{z})[\varphi(\hat{z}) - \varphi(\hat{z}_{\tau_j})] = I_1(\varphi) + I_2(\varphi),$$

where

$$I_1(\varphi) \equiv \int_{\hat{O}_{\epsilon,\lambda_1,\lambda_2}^{n,j}} dx_1 \cdots d\mathbf{x}_j \cdots dx_n \int_{y_j^0}^{y_j'^0} dy_j''^0 \underline{\mathfrak{W}}^T(\hat{x} - i\hat{y}'')$$

$$\times [\varphi(x - iy'') - \varphi(x_{\tau_j} - iy_{\tau_j}'')]\big|_{x_j^0 = x_{j+1}^0 - \lambda_1 \|\mathbf{x}_j - \mathbf{x}_{j+1}\|}$$

$$I_2(\varphi) \equiv \int_{\hat{O}_{\epsilon,\lambda_1,\lambda_2}^{n,j}} dx_1 \cdots d\mathbf{x}_j \cdots dx_n \int_{y_j'^0}^{y_j^0} dy_j''^0 \underline{\mathfrak{W}}^T(\hat{x} - i\hat{y}'')$$

$$\times [\varphi(\hat{x} - i\hat{y}'') - \varphi(\hat{x}_{\tau_j} - i\hat{y}_{\tau_j}'')]_{x_j^0 = x_{j+1}^0 + \lambda_1 \|\mathbf{x}_j - \mathbf{x}_{j+1}\|},$$

$$y_r'' \equiv y_r \equiv y_r' \text{ for } r \neq j, \quad y_j'' \equiv \mathbf{y}_j \equiv \mathbf{y}_j'.$$

Here we omitted the additional term

$$\int_{\hat{O}_{\epsilon,\lambda_1,\lambda_2}^{n,j}} dx_1 \cdots d\mathbf{x}_j \cdots dx_n \int_{x_{j+1}^0 - \lambda_1 \|\mathbf{x}_j - \mathbf{x}_{j+1}\|}^{x_{j+1}^0 + \lambda_1 \|\mathbf{x}_j - \mathbf{x}_{j+1}\|} dx_j^0 \underline{\mathfrak{W}}^T(\hat{x} - i\hat{y}')$$

$$\times [\varphi(x - iy') - \varphi(x_{\tau_j} - iy_{\tau_j}')]$$

$$= \int_{O_{\epsilon,\lambda_1,\lambda_2}^{n,j}} d\hat{x}[\underline{\mathfrak{W}}^T(\hat{x} - i\hat{y}') - \underline{\mathfrak{W}}^T(\hat{x}_{\tau_j} - i\hat{y}_{\tau_j}')]\varphi(\hat{x} - i\hat{y}'),$$

which is zero, due to the Wightman property. By (4.2) and (4.5) both $|I_1(\varphi)|$ and $|I_2(\varphi)|$ have an upper bound independent of $\varphi \in S$; hence we have finally proved the desired continuity property:

For arbitrary $\epsilon, \lambda_1, \lambda_2 > 0$ with $\lambda_1 < 1$ and for arbitrary $n, j \in \mathbb{Z}_+$ with $0 < j < n$ (4.1) holds for every subset S of $S^1(\mathbb{R}^{4n})$ fulfilling (4.2) and (4.3).

As in Ref. 11, let us now introduce the following notation:

$$\underline{\mathfrak{W}}_{\tau}^T(\hat{x}) \equiv \underline{\mathfrak{W}}^T(x_{\tau(1)}, \ldots, x_{\tau(n)}),$$

$$\hat{a}_0 \equiv ((0, \mathbf{a}_1), \ldots, (0, \mathbf{a}_n)),$$

$$m_{\tau, \hat{a}_0} \equiv \min_{(r,r') \in \{1, \ldots, n'\} \times \{n'+1, \ldots, n\}} \|\mathbf{a}_{\tau(r)} - \mathbf{a}_{\tau(r')}\|,$$

$$\mathfrak{F}_\varphi^\tau(\hat{a}_0) \equiv \int d\hat{x}\, \underline{\mathfrak{W}}_\tau^T(\hat{x})\varphi(\hat{x} - \hat{a}_0),$$

$$S_{n,n'} \equiv \{\pi \in S_n: \pi(r) < \pi(r') \text{ for } 0 < r < r' \leq n'$$

$$\text{as well as for } n' < r < r' \leq n\}.$$

Then the following lemma, which was derived from essential locality in Ref. 11, may still be derived from the continuity property just established:

Lemma 4.2: Let N be a nonnegative integer and let $\pi \in S_{n,n'}$ and $\varphi \in S^1(\mathbb{R}^{4n})$. Then there is a constant C for which the inequality

$$m_{\tau, \hat{a}_0}^N \left| \mathfrak{F}_\varphi^1(\hat{a}_0) - \mathfrak{F}_\varphi^\tau(\hat{a}_0) \right| < C$$

holds for arbitrary $\mathbf{a}_1, \ldots, \mathbf{a}_n \in \mathbb{R}^3$ with $\|\mathbf{a}_r - \mathbf{a}_{r'}\| \leq (n-1)m_{\tau, \hat{a}_0}$.

Proof: Recall that $\underline{\mathfrak{W}}^T(\hat{x}) - \underline{\mathfrak{W}}_s^T(x)$ can be written as a finite sum of generalized functions of the form

$$\underline{\mathfrak{W}}^T(x_{\tau'(1)}, \ldots, x_{\tau'(n)})$$

$$- \underline{\mathfrak{W}}^T(x_{\tau'(1)}, \ldots, x_{\tau'(j+1)}, x_{\tau'(j)}, \ldots, x_{\tau'(n)})$$

with suitable $j \in \{1, \ldots, n-1\}$ and suitable $\pi' \in S_n$ fulfilling $\pi'(j) \in \{\pi(n'+1), \ldots, \pi(n)\}$ and $\pi'(j+1) \in \{\pi(1), \ldots, \pi(n')\}$. Hence, by the continuity property established above,

$$\sup_{\varphi \in S} \left| \int d\hat{x}\, [\underline{\mathfrak{W}}^T(\hat{x}) - \underline{\mathfrak{W}}_\tau^T(\hat{x})]\varphi(\hat{x}) \right| < \infty$$

holds for every set $S \subset S^1(\mathbb{R}^{4n})$ and $\epsilon > 0$ for which the conditions (4.3) and

$$\sup_{\varphi \in S} \sup_{\hat{z} \in U_\epsilon^c(M_\epsilon^\tau)} \sup_{\hat{\alpha} \in \mathbb{Z}_+^{4n}} \epsilon^{|\hat{\alpha}|} \hat{\alpha}^{-\hat{\alpha}} \|z\|^N \left| \varphi^{(\hat{\alpha})}(\hat{z}) \right|$$

$$< \infty \text{ for } N \in \mathbb{Z}_+ \tag{4.6}$$

are fulfilled. Here we define $\hat{x}$ to be in M_ϵ^τ iff for arbitrary $(j,j') \in \{1, \ldots, n'\} \times \{n'+1, \ldots, n\}$ at least one of the inequalities

$$\left| x_{\tau(j)}^0 - x_{\tau(j')}^0 \right| \geq \tfrac{1}{2} \|\mathbf{x}_{\tau(j)} - \mathbf{x}_{\tau(j')}\|,$$

$$\|\mathbf{x}_{\tau(j)} - \mathbf{x}_{\tau(j')}\| \leq \epsilon,$$

or

$$\|\mathbf{x}_{\tau(j)} - \mathbf{x}_{\tau(j')}\| \leq (1/n)\|\mathbf{x}_{\tau(r)} - \mathbf{x}_{\tau(r')}\|$$

for suitable $r, r' \in \{1, \ldots, n\}$

holds. Let N' be an arbitrary nonnegative integer. Then, by translation invariance, our proof is complete if we show (4.3) and (4.6) for the special set

$$S = \{\varphi_{\hat{a}_0}^z(\hat{x}) \equiv m_{\tau, \hat{a}_0}^{N'} \varphi(\hat{x} - \hat{a}_0): \mathbf{a}_2, \ldots, \mathbf{a}_n \in \mathbb{R}^3, \mathbf{a}_1 \equiv 0\}$$

and for suitable $\epsilon > 0$. While (4.3) is a simple consequence of the definition of $S^1(\mathbb{R}^{4n})$ and Lemma 3.1, condition (4.6) is (by Lemma 3.1, again) a consequence of

$$\sup_{\varphi \in S} \sup_{\hat{x} \in M_1^\tau} \sup_{\hat{\alpha} \in \mathbb{Z}_+^{4n}} \epsilon^{|\hat{\alpha}|} \hat{\alpha}^{-\hat{\alpha}} \|\hat{x}\|^N \left| \varphi^{(\hat{\alpha})}(\hat{x}) \right|$$

$$< \infty \text{ for } N \in \mathbb{Z}_+. \tag{4.7}$$

So we are left to prove (4.7):

Let $\hat{x} + \hat{a}_0 \in M_1^\tau$. Then, for suitable $(j,j') \in \{1, \ldots, n'\}$

$\times \{n' + 1, \ldots, n\}$, we have

$$\|\mathbf{a}_{\tau(j)} - \mathbf{a}_{\tau(j')}\| = m_{\tau,\hat{a}_0}$$

and at least one of the inequalities

$$\left| x^0_{\tau(j)} - x^0_{\tau(j')} \right| \geq \tfrac{1}{2} \|(\mathbf{x}_{\tau(j)} - \mathbf{x}_{\tau(j')}) + (\mathbf{a}_{\tau(j)} - \mathbf{a}_{\tau(j')})\|,$$

$$\|(\mathbf{x}_{\tau(j)} - \mathbf{x}_{\tau(j')}) + (\mathbf{a}_{\tau(j)} - \mathbf{a}_{\tau(j')})\| \leq 1,$$

or

$$\|(\mathbf{x}_{\tau(j)} - \mathbf{x}_{\tau(j')}) + (\mathbf{a}_{\tau(j)} - \mathbf{a}_{\tau(j')})\|$$

$$\leq \tfrac{1}{n} (\|x_{\tau(r)} - x_{\tau(r')}\| + \|\mathbf{a}_{\tau(r)} + \mathbf{a}_{\tau(r')}\|)$$

for suitable r, r' holds. In any case we have the inequality $m_{\tau,\hat{a}_0} \leq 1 + 2(n+1)\|\hat{x}\|$, which allows the following estimate:

$$\sup_{\hat{x} \in M^{\tau}_1} \|\hat{x}\|^N \left| \varphi^{(\hat{\alpha})}_{\hat{a}_0}(\hat{x}) \right|$$

$$\leq m^{N'}_{\tau,\hat{a}_0} \sup_{\hat{x}+\hat{a}_0 \in M^{\tau}_1} (\|\hat{x}\| + (n-1)^2 m_{\tau,\hat{a}_0})^N \left| \varphi^{(\hat{\alpha})}(\hat{x}) \right|$$

$$\leq \sup_{\hat{x} \in \mathbb{R}^{4n}} (2(n+1)^3 \|\hat{x}\| + (n-1)^2)^{N+N'} \left| \varphi^{(\hat{\alpha})}(\hat{x}) \right|.$$

This, together with the definition of $S^1(\mathbb{R}^{4n})$, implies (4.7). ∎

As shown in Ref. 11, following Ruelle[10] one can derive from Lemma 4.2 and the mass gap condition the following:

Cluster property: Let N be a nonnegative integer and let φ be a test function from $S^1(\mathbb{R}^{4n})$. Then there is a constant C such that the inequality

$$\|\hat{a}_0\|^N \left| \mathfrak{F}^1_\varphi(\hat{a}_0) \right| < C$$

holds for $\mathbf{a}_1 = 0$ and for arbitrary $\mathbf{a}_2, \ldots, \mathbf{a}_n \in \mathbb{R}^3$.

ACKNOWLEDGMENT

[1]This was first proved by A.S. Wightman, Ann. Inst. Henri Poincaré **1**, 403 (1964); for a simplified proof see also A. Jaffe, Phys. Rev. **158**, 1454 (1967).

[2]For a detailed discussion on this point see A.S. Wightman and L. Gårding, Arkiv Fysik **28**, 129 (1964).

[3]The standard textbooks on this subject are: R.F. Streater and A.S. Wightman, *PCT, Spin and Statistics, and All That* (Benjamin, New York, 1964); R. Jost, *The General Theory of Quantized Fields* (Am. Math. Soc., Providence, R.I., 1965).

[4]A. Jaffe, Ref. 1; a simplified treatment is given by W. Lücke, Int. J. Theor. Phys. **7**, 231 (1973); for an elegant fundamental exposition of the corresponding mathematical structure see C. Roumieu, J. Analyse Math. **10**, 153 (1962–63); a more elaborate exposition with emphasis on differential equations is given by G. Björck, Arkiv Mat. **6**, 351 (1966).

[5]I.M. Gelfand and G.E. Schilow, *Verallgemeinerte Funktionen*, Vol. II (VEB Deutsch. Verl. Wissensch., Berlin, 1962), Chap. IV.

[6]F. Constantinescu, J. Math. Phys. **12**, 193 (1971).

[7]C. Roumieu, Ann. Sci. École Norm. Supl. (3) **77**, 41 (1960), Chap. II, Theorem 1.

[8]See A. Rieckers, Int. J. Theor. Phys. **4**, 55 (1971); in view of Sec. 3 of the present paper see also W. Lücke, J. Phys. A **7**, 2258 (1974).

[9]K. Osterwalder and R. Schrader, Commun. Math. Phys. **31**, 83 (1973), and **42**, 281 (1975); in particular see V. Glaser, Commun. Math. Phys. **37**, 257 (1974).

[10]D. Ruelle, Helv. Phys. Acta **35**, 147 (1962).

[11]J. Bümmerstede and W. Lücke, Commun. Math. Phys. **37**, 121 (1974).

[12]Just choose $M = R^n$ in Lemma 2 of Ref. 11.

[13]L. Gårding and J. Lions, Nuovo Cimento Suppl. **14**, 9 (1959), Section 12.

[14]See Ref. 8, Theorem A1.

[15]The notation $^{(T)}$ means that the index T may or may not be present in the relation under consideration.

[16]See Ref. 8, Theorem A2; more general bounds may be established following a method by G.R. Screaton and A. Truman, J. Math. Phys. **14**, 982 (1973).

[17]See Ref. 6, Theorem 2; note that this theorem is a consequence of Lemmas 3.1 and 3.2 of the present paper.

[18]See R. Jost, Ref. 3, p. 83.

[19]For the completely nonlocalizable case see J. Bümmerstede and W. Lücke, J. Math. Phys. **16**, 1203 (1975).

[20]We use standard notation:

$$\mathbb{Z}_+ = \{0, 1, 2, \cdots\}, \quad \mathbb{Z}^n_+ = \mathbb{Z}_+ \times \cdots \times \mathbb{Z}_+, \quad |\alpha| \equiv \alpha^1 + \cdots + \alpha^n,$$

$$\sigma^\alpha \equiv \prod_{j=1}^n (\sigma^j)^{(\alpha^j)}, \quad \alpha! \equiv \prod_{j=1}^n \alpha^j!, \quad \|\sigma\| \equiv (\sum_{j=1}^n |\sigma^j|^2)^{1/2}.$$

[21]We write

$$U^c_\epsilon(L) \equiv \{\sigma \in \mathbb{C}^n : \|\sigma - \chi'\| < \epsilon \text{ for at least one } \chi' \in L\} \quad \text{and}$$

$$U_\epsilon(L) \equiv \{\chi \in \mathbb{R}^n : \|\chi - \chi'\| < \epsilon \text{ for at least one } \chi' \in L\} \text{ for}$$

arbitary $L \subset \mathbb{R}^n$; and $\epsilon > 0$.

[22]I.M. Gelfand and G.E. Schilow, *Verallgemeinerte Funktionen*, Vol. I (VEB Deutsch. Verl. Wissensch., Berlin, 1967), Chap. I, §2, No. 5.

[23]We denote by S_n the set of all permutations of $1, \ldots, n$.

[24]See Ref. 5, p. 180.

[25]See R. Jost, Ref. 3, Appendix II.

Representations of the universal covering group of SU(1,1) and their bilinear and trilinear invariant forms

W. Rühl and B. C. Yunn

Fachbereich Physik der Universität Kaiserslautern, 6750 Kaiserslautern, Germany
(Received 15 September 1975)

The unitary irreducible and many nonunitary representations of the universal covering group of SU(1,1) are given in a realization on certain spaces of functions. We discuss intertwining operators for these representations and their connection with the discrete series. The tensor product decomposition is performed by means of an integral transformation. A completeness relation for these integral kernels is derived.

1. INTRODUCTION

Our study of the universal convering group G of SU(1,1) is motivated by the fact that conformal invariant quantum field theories in two-dimensional space—time (like the Thirring model) exhibit the symmetry group $G \otimes G$, which is the universal covering group of the conformal group for two-dimensional Minkowski space. Representations of the group $G \otimes G$ are used in the construction of such models. It is hoped that by means of harmonic analysis on the symmetry group $G \otimes G$ one gets insight into certain features of these models that can be carried over to quantum field theories in four-dimensional Minkowski space.

Representations of the group G, namely those of the discrete series, have been given by Bargmann.[1] Both the principal and the discrete series have been presented in infinitesimal form (as matrices) and in global form (as operators in function spaces) by Pukanszky.[2] This author gave the supplementary series in infinitesimal form in addition. His main result was the proof of the Plancherel formula.

In this article we present all unitary irreducible represenations and also a great many nonunitary ones in an operator realization on a function space. We construct bilinear invariant forms and their kernels that are related with the intertwining operators. For the supplementary series such kernels supply us with an invariant scalar product. Finally we study trilinear invariant forms and their kernels that are related with the problem of the tensor product decomposition of two representations. We derive a completeness relation for these kernels. This relation is equivalent with the completeness relation for the Wigner coefficients of G. These Wigner coefficients can be defined as the trilinear invariant forms for the elements of the canonical basis. But we do not introduce them explicitly, mainly because this would involve reasoning for a specific choice of normalization which is of no use for us later.

We emphasize that the coordinate functions, that is, the matrix elements of the representation operators in the canonical basis (the c or d functions), and the multilinear invariant forms for the elements of the canonical basis can be obtained from the corresponding expressions for the group SU(1,1) by a proper interpolation and analytic continuation in the helicities. For the representations of SU(1,1) we refer to the textbooks,

Refs. 3,4. Our work profited most from the work of Ferretti and Verde[5] and Wang[6] who deal with the tensor product decomposition for SU(1.1). In fact, there exists older work on this subject by Andrews and Gunson[7] and Holman and Biedenharn.[8] Andrews and Gunson were the first authors to apply the Burchnall—Chaundy expansion in this context. This method has been used later by Wang and has also been adopted in this article. It seems to us much more elegant than the formalism of Holman and Biedenharn which is based on interpolating and analytically continuing the respective formulae of the group SU(2). The proofs given in Refs. 5, 6 for the deformation of a contour in the complex j-plane are reduced to a lemma that is given in the Appendix.

2. REPRESENTATIONS

We denote elements of SU(1,1) by

$$v = \begin{pmatrix} \alpha & \beta \\ \bar{\beta} & \bar{\alpha} \end{pmatrix}, \quad \det v = 1. \tag{1}$$

An element of G can be characterized by the matrix v and a real number ξ, $-\infty < \xi < +\infty$, $\xi = \arg \alpha$. Thus the group G possesses an infinite number of sheets over SU(1,1).

Next we introduce the space $(0 \leqslant \tau < 1)$

$$D_\tau = \{g(\varphi) \in C^\infty(-\infty, +\infty) \mid g(\varphi + 2\pi) = \exp(2\pi i \tau) g(\varphi)\} \tag{2}$$

of complex valued functions. Extracting a factor

$$g(\varphi) = \exp(i\tau\varphi) f(\varphi), \quad g \in D_\tau \tag{3}$$

yields a periodic function $f(\varphi)$ that possesses a rapidly converging Fourier series

$$f(\varphi) = \sum_{m=-\infty}^{+\infty} a_m \exp(im\varphi). \tag{4}$$

By this property the elements of D_τ can be characterized.

In D_τ we introduce the canonical basis

$$\{g_q(\varphi) \mid g_q(\varphi) = \exp(iq\varphi), \quad q = \tau + m, \; m \text{ integral},$$
$$-\infty < m < +\infty\} \tag{5}$$

and the canonical norm

$$\|g\|^2 = \frac{1}{2\pi} \int_0^{2\pi} |g(\varphi)|^2 \, d\varphi. \tag{6}$$

Completing the space D_τ with this norm (6) yields a Hilbert space L_τ^2.

The space D_τ can be made to carry a representation of G, denoted by the symbol $\chi = (j, \tau)$, in the following way. Let us denote now

$$v^{-1} = \begin{pmatrix} \alpha & \beta \\ \bar\beta & \bar\alpha \end{pmatrix}, \quad v = \begin{pmatrix} \bar\alpha & -\beta \\ -\bar\beta & \alpha \end{pmatrix}. \tag{7}$$

For arbitrary $j \in \mathbb{C}_1$ we define the linear operator in D_τ

$$T_{v,\xi}^\chi g(\varphi) = |\alpha + \beta \exp(-i\varphi)|^{2j-1} g(\varphi_{v,\xi}), \tag{8}$$

where

$$\exp(i\varphi_{v,\xi}) = \frac{\alpha + \beta \exp(-i\varphi)}{\bar\alpha + \bar\beta \exp(+i\varphi)} \exp(i\varphi) \tag{9}$$

is independent of ξ, but $\varphi_{v,\xi}$ is not

$$\varphi_{v,\xi} = \varphi + 2 \arg(\alpha + \beta \exp(-i\varphi)). \tag{10}$$

Because $|\alpha| > |\beta|$, the rhs of (10) is uniquely defined by $\xi = \arg\alpha$, namely

$$|\arg(\alpha + \beta \exp(-i\varphi)) - \arg\alpha| < \pi/2. \tag{11}$$

It is a straightforward task to show that these operators satisfy the group multiplication law and present bounded operators from D_τ into D_τ with respect to the canonical norm (6). Continuity in the group element can also be proven with respect to the same norm. Similar assertions are true in other topologies like the D_{L^2}-topology.[9]

In the canonical basis the operators $T_{v,\xi}^\chi$ possess the matrix elements ("coordinate functions")

$$T_{v,\xi}^\chi g_{q_2}(\varphi) = \sum_{q_1} C_{q_1 q_2}^\chi(v, \xi) g_{q_1}(\varphi) \tag{12}$$

with the integral representation

$$C_{q_1 q_2}^\chi(v, \xi) = \frac{1}{2\pi} \int_0^{2\pi} \exp(-i(q_1 - q_2))\varphi$$

$$\times (\alpha + \beta \exp(-i\varphi))^{-1/2+j+q_2}(\bar\alpha + \bar\beta \exp(+i\varphi))^{-1/2+j-q_2} d\varphi. \tag{13}$$

We define coordinates on $SU(1,1)$ by

$$v^{-1} = \exp(i\sigma_3 \varphi_2/2) \exp(\sigma_1 \eta/2) \exp(i\sigma_3 \varphi_1/2), \quad \eta \ge 0. \tag{14}$$

Then (13) can be evaluated

$$C_{q_1 q_2}^\chi(v, \xi) = \exp(i\varphi_1 q_1 + i\varphi_2 q_2) c_{q_1 q_2}^j (\cosh\eta), \tag{15a}$$

$$c_{q_1 q_2}^j(u)$$
$$= \begin{pmatrix} -\frac{1}{2} + j - q_2 \\ q_1 - q_2 \end{pmatrix} \left(\frac{1+u}{2}\right)^{-(q_1+q_2)/2} \left(\frac{u-1}{2}\right)^{(q_1-q_2)/2}$$
$$\times {}_2F_1(\tfrac{1}{2} + j - q_2, \tfrac{1}{2} - j - q_2; q_1 - q_2 + 1; \tfrac{1}{2}(1-u)),$$
$$q_1 \ge q_2, \tag{15b}$$

$$c_{-q_1,-q_2}^j(u) = \frac{\Gamma(\tfrac{1}{2} + j + q_2)\Gamma(\tfrac{1}{2} + j - q_2)}{\Gamma(\tfrac{1}{2} + j + q_1)\Gamma(\tfrac{1}{2} + j - q_1)}$$
$$\times c_{q_2 q_1}^j(u), \quad q_1 \le q_2. \tag{15c}$$

The principal series is obtained if j is purely imaginary. In this case the canonical norm (6) can be shown to be invariant. Thus the principal series can be realized in the Hilbert spaces L_τ^2 with the invariant scalar product

$$(g_1, g_2) = \frac{1}{2\pi} \int_0^{2\pi} \overline{g_1(\varphi)} g_2(\varphi) d\varphi. \tag{16}$$

The discrete series is obtained as follows. We set

$$g(\varphi) = \exp(ik\varphi) h(\varphi) \tag{17}$$

and have

$$T_{v,\xi}^\chi h(\varphi) = \exp(-ik(\varphi - \varphi_{v,\xi})) |\alpha + \beta \exp(-i\varphi)|^{2j-1}$$
$$\times h(\varphi_{v,\xi}). \tag{18}$$

For $k = \pm(j - \tfrac{1}{2})$ the multiplier in (18) is $(\alpha + \beta \exp(-i\varphi))^{2j-1}$, respectively $(\bar\alpha + \bar\beta \exp(i\varphi))^{2j-1}$, due to (10). These multipliers are boundary values of the anti-holomorphic function $(\alpha + \beta\bar z)^{2j-1}$, respectively the holomorphic function $(\bar\alpha + \bar\beta z)^{2j-1}$ in the unit circle $|z| < 1$. If h is periodic, viz.,

$$k \overset{\simeq}{=} \tau \mod 1, \tag{19}$$

then the negative, respectively positive, frequency parts [by negative frequency part we mean that the coefficients a_m in the Fourier series (4) of $h(\varphi)$ vanish for $m > 0$; the positive frequency part is defined correspondingly] constitute invariant subspaces. Thus D_τ possesses an invariant subspace $\mathcal{J}_\chi^{(-)}$ (respectively $\mathcal{J}_\chi^{(+)}$) under $T_{v,\xi}^\chi$ provided

$$j - \tfrac{1}{2} \overset{\simeq}{=} \tau \mod 1 \tag{20}$$

(respectively $j - \tfrac{1}{2} \overset{\simeq}{=} -\tau \mod 1$).

These subspaces are spanned by the canonical basis elements

$$\mathcal{J}_\chi^{(-)}: q = j - \tfrac{1}{2} - m, \quad m = 0, 1, 2, \cdots \tag{21}$$

(respectively $\mathcal{J}_\chi^{(+)}: q = -j + \tfrac{1}{2} + m, \quad m = 0, 1, 2, \cdots$).

If (19) is fulfilled, then the positive (negative) frequency part of $h(\varphi)$ can be continued into the unit circle $|z| < 1$ to yield a holomorphic (antiholomorphic) function $\tilde h^{(+)}(z)$ $[\tilde h^{(-)}(z)]$ of z. Following from the subspaces $\mathcal{J}_\chi^{(\pm)}$ of D_τ, we have obtained certain linear spaces of such holomorphic or antiholomorphic functions. The scalar product for the discrete series of $SU(1,1)$[1] can be taken over unaltered,

$$(g_1, g_2)_{\pm,n} = \frac{2n-1}{\pi} \int_{|z|<1} \tilde h_1^{(\pm)}(z) \tilde h_2^{(\pm)}(z)$$
$$\times (1 - |z|^2)^{2n-2} dx\, dy, \tag{22}$$

$$n = \tfrac{1}{2} - j \text{ real.}$$

This integral converges obviously for $n > \tfrac{1}{2}$. But the scalar product can be extended to all $n > 0$, as we shall see in a moment. The invariance of the scalar product (22) is easy to prove. Completing the linear spaces by means of (22) or its extension to $n > 0$ leads to the unitary representations of the discrete series. Namely, the positive frequency parts yield the positive discrete series and the negative frequency parts the negative discrete series.

The coordinate functions for the discrete series come out as [using the functions (15b, c) as analytic expressions in j]

$$c_{q_1 q_2}^{j, \text{discr}}(u) = N_{|q_1|}^{-1} N_{|q_2|} \, c_{q_1 q_2}^{j}(u),$$ (23)

where for the positive (negative) discrete series

$$q_{1,2} = -j + \tfrac{1}{2} + m,$$

$$(q_{1,2} = +j - \tfrac{1}{2} - m),$$ (24)

$$m = 0, 1, 2, \cdots, \quad j < +\tfrac{1}{2}$$

and N_q denotes

$$N_q = \frac{\Gamma(\tfrac{1}{2} - j + q)}{\Gamma(1 - 2j)\Gamma(\tfrac{1}{2} + j + q)}^{1/2}.$$ (25)

N_q^{-1} is the norm of the basis element $g_q(\varphi)$ due to the scalar product (22). This expression (25) can in turn be used to define the scalar product in a basis dependent fashion (which makes it more difficult to prove invariance). It shows that a positive scalar product can in fact be defined for all real j smaller than $\tfrac{1}{2}$ corresponding to $n > 0$, as was asserted above. The negative discrete series can be treated in an analogous fashion.

3. THE INTERTWINING OPERATOR

We define an operator S on D_τ for all j in the left half j plane by the convolution integral

$$Sg(\varphi_1) = \int_0^{2\pi} S(\chi^c, \varphi_1 | \chi, \varphi_2) g(\varphi_2) \, d\varphi_2$$ (26a)

$$= N(\chi) \int_0^{2\pi} \left| 2 \sin\frac{\varphi_1 - \varphi_2}{2} \right|^{-2j-1}$$

$$\times \exp(+2i\tau\sigma(\varphi_1 - \varphi_2)) g(\varphi_2) \, d\varphi_2.$$ (26b)

with

$$\sigma(\varphi) = \begin{cases} \arg(\sin(\varphi/2 - i0)) + \pi/2 \\ \pi/2 \quad \text{for } 0 < \varphi < 2\pi. \end{cases}$$ (27)

The function $\sigma(\varphi)$ satisfies

$$\sigma(\varphi + 2\pi) = \sigma(\varphi) + \pi,$$ (28)

$$\sigma((\varphi_1)_{v,t} - (\varphi_2)_{v,t}) = \sigma(\varphi_1 - \varphi_2).$$ (29)

We choose the normalization constant $N(j, \tau)$ for convenience as

$$N(\chi)^{-1} = N(j, \tau)^{-1} = \Gamma(-2j) \, 2 \sin\pi(\tfrac{1}{2} + j - \tau).$$ (30)

The operator S turns out to be diagonal on the canonical basis

$$Sg_q(\varphi) = S_{qq} g_q(\varphi),$$ (31a)

$$S_{qq} = S_{qq}(j) = \Gamma(\tfrac{1}{2} + j - q)/\Gamma(\tfrac{1}{2} - j - q).$$ (31b)

From (31b) we see that the definition of the operator S can be extended to all complex j except a discrete set of singular values on the real axis where some S_{qq} become infinite. We want to include in the set of singular points also those where S_{qq} vanishes. Of course, the whole set depends on the choice of the normalization constant (30). This extended operator is called the intertwining operator. The singular points are connected with the discrete series and will be investigated in a moment.

The name of the operator S results from the property that it intertwines two representations

$$\chi = (j, \tau) \quad \text{and} \quad \chi^c = (-j, \tau)$$

in the sense that

$$ST_{v,t}^\chi = T_{v,t}^{\chi^c} S$$ (32)

on D_τ provided only that we are not at a singular point. In this case S is a mapping of D_τ onto D_τ and possesses an inverse S^{-1}. We call χ^c the "conjugate" representation of χ. If j is purely imaginary, S can be continued to a unitary operator on L_τ^2 so that two representations χ, χ^c of the principal series turn out to be equivalent.

We denote the representation χ^d the "dual" representation of χ if

$$\chi = (j, \tau), \quad \chi^d = (-j, -\tau).$$

Let $g_1 \in D_{-\tau}$, $g_2 \in D_\tau$. Then

$$B(g_1, g_2) = \int_0^{2\pi} g_1(\varphi) g_2(\varphi) \, d\varphi$$ (33)

is an invariant bilinear form if g_1 transforms as χ and g_2 as its dual χ^d. This assertion follows immediately from (2), (8), and

$$\frac{d\varphi_{v,t}}{d\varphi} = |\alpha + \beta \exp(-i\varphi)|^{-2}.$$ (34)

If $g_1 \in D_{-\tau}$, $g_2 \in D_\tau$, then the bilinear form

$$B(g_1, g_2) = \int_0^{2\pi} g_1(\varphi)(Sg_2)(\varphi) \, d\varphi$$ (35)

is invariant, provided that g_1 transforms as $\chi^{d,c}$ if g_2 transforms as χ. We shall make use of this fact in the construction of the invariant scalar product for the supplementary series.

Let $\tfrac{1}{2} - j - \tau$ be an integer, but $\tfrac{1}{2} + j - \tau$ not. This implies that all matrix elements S_{qq} are finite, and some are zero. We can also write this assumption as [see (22), (23)]

$$\tfrac{1}{2} - j = n, \quad n \cong + \tau \mod 1, \quad 2j \text{ not integral.}$$ (36)

Then the intertwining operator S annihilates the subspace spanned by the vectors $g_q(\varphi)$ with

$$q = n + m, \quad m = 0, 1, 2, \cdots,$$

which we denote $\mathcal{J}_\chi^{(+)}$. This null space $\mathcal{J}_\chi^{(+)}$ of S is an invariant subspace and carries a positive discrete series representation as we know from Sec. 2. Similarly we have an invariant subspace $\mathcal{J}_\chi^{(-)}$ if $n \cong -\tau \mod 1$, spanned by the elements $g_q(\varphi)$ with $q = -n - m$, $m = 0, 1, 2, \cdots$.

We consider two functions $g(\varphi)$, $g_2(\varphi)$ of D_τ which do not have any components in $\mathcal{J}_\chi^{(+)}$. We set $g_1(\varphi) = \overline{g(\varphi)}$ in (35). Since j is real, g transforms such as g_2. Consequently, (35) defines a sesquilinear invariant form on $D_\tau / \mathcal{J}_\chi^{(+)}$. After a renormalization of S we get a sesquilinear form on $\mathcal{J}_\chi^{(+)}$ itself. Namely, replace the rhs of (30) by $\Gamma(-2j) \, 2 \sin\pi(\tfrac{1}{2} + j + \tau)$ and denote the new operator by S'. Instead of (31b) we obtain

$$S'_{qq} = \Gamma(\tfrac{1}{2} + j + q)/\Gamma(\tfrac{1}{2} - j + q).$$ (37)

S' does not annihilate $\mathcal{J}_\chi^{(+)}$ but $\mathcal{J}_\chi^{(-)}$. It gives on $\mathcal{J}_\chi^{(+)}$

$$S'_{qq} = \Gamma(1 - 2j)^{-1} N_q^{-2}$$ (38)

so that, up to a constant factor, S' generates the scalar product (22) of the discrete series.

1523   J. Math. Phys., Vol. 17, No. 8, August 1976

W. Rühl and B.C. Yunn   1523
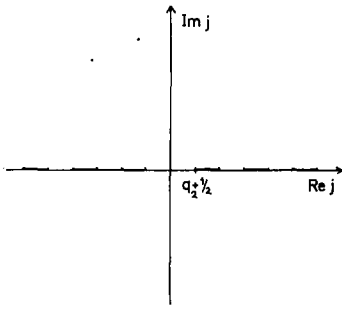
FIG. 1. The cuts of the function $\{j\}^{1/2}$ in the j plane for the case $q_1 \geq q_2 \geq 0$.

One can elaborate further on the singular points of the intertwining operator, in particular on the case that both $\frac{1}{2} \pm j - \tau$ are integral. In this case $2j$ is also integral. It results in a discrete series representation of $SU(1,1)$. It is only in this case that subspaces of the type $\mathcal{E}_\chi$ appear as intersections of $\mathcal{J}_\chi^{(+)}$ and $\mathcal{J}_\chi^{(-)}$ that carry finite-dimensional "spinorial" representations.

The bilinear form (35) leads to an invariant scalar product of the form

$$(g_1, g_2) = \int_0^{2\pi} \overline{g_1(\varphi)}\, (Sg_2)(\varphi)\, d\varphi \tag{39}$$

also for the supplementary series (after an eventual adjustment of the sign). We need only require that j is real and that

$$\frac{S_{qq}}{S_{q-1,q-1}} = \frac{\frac{1}{2} - j - q}{\frac{1}{2} + j - q} > 0 \tag{40}$$

for all q. We require $j < 0$ in order to render the integral (39) convergent for $g_{1,2} \in \mathcal{D}_\tau$. All points in the domain

$$-\left|\tfrac{1}{2} - \tau\right| < j < 0, \quad 0 \leqslant \tau < 1, \tag{41}$$

and only these solve these constraints, as has first been found by Pukanszky.[2]

Finally use is made of the intertwining operator in defining new coordinate functions, the "d functions"

$$D_{q_1 q_2}^\chi(v, \xi) = \exp(iq_1\varphi_1 + iq_2\varphi_2)\, d_{q_1 q_2}^j(\cosh\eta), \tag{42a}$$

$$d_{q_1 q_2}^j(u) = S'^{1/2}_{q_1 q_1}(j)\, S'^{-1/2}_{q_2 q_2}(j)\, c_{q_1 q_2}^j(u). \tag{42b}$$

The definition of the square root is a matter of convenience. We set

$$\frac{\Gamma(\frac{1}{2} + j - q_2)}{\Gamma(\frac{1}{2} + j - q_1)}\, S'^{1/2}_{q_1 q_1}(j)\, S'^{-1/2}_{q_2 q_2}(j)$$

$$= \left\{\frac{\Gamma(\frac{1}{2} - j + q_1)\Gamma(\frac{1}{2} + j + q_1)}{\Gamma(\frac{1}{2} - j + q_2)\Gamma(\frac{1}{2} + j + q_2)}\right\}^{1/2} = \{j\}^{1/2}, \tag{43}$$

where the right-hand side is defined as a symmetric ($j \to -j$) analytic function of j. For $q_1 \geqslant q_2$ $\{j\}$ is a polynomial of j of degree $2(q_1 - q_2)$. We define the cuts of $\{j\}^{1/2}$ as in Fig. 1. In particular a cut extends to ∞ if and only if $q_1 - q_2$ is odd. On the imaginary axis we choose positive real values. Correspondingly the asymptotic behavior of $\{j\}^{1/2}$ is

$$\{j\}^{1/2} \underset{|j| \to \infty}{\sim} (\mp ij)^{q_1 - q_2}(1 + O(\tfrac{1}{j})), \quad 0 < \pm \arg j < \pi. \tag{44}$$

From the definition several symmetries of the d functions can be deduced. From (32) or explicitly from (15a) we have

$$c_{q_1 q_2}^{-j}(u) = S'_{q_1 q_1}(j)\, c_{q_1 q_2}^j(u)\, S'^{-1}_{q_2 q_2}(j). \tag{45}$$

For the d functions this relation simplifies to

$$d_{q_1 q_2}^{-j}(u) = d_{q_1 q_2}^{+j}(u). \tag{46}$$

Moreover, we find from (15c)

$$d_{q_1 q_2}^j(u) = (-1)^{q_1 - q_2} d_{q_2 q_1}^j(u) \tag{47}$$

$$= (-1)^{q_1 - q_2} d_{-q_1, -q_2}^j(u). \tag{48}$$

The relation (48) connects the representations χ and $\chi^{c,d}$. These form a pair due to the automorphism of the group G

$$v \to v^{-1,T}, \quad \xi \to -\xi$$

With (48) we can easily derive

$$D_{q_1 q_2}^\chi(v^{-1,T}, -\xi) = D_{-q_1, -q_2}^{\chi^{c,d}}(v, \xi). \tag{49}$$

If $\mathcal{D}_\tau$ possesses an invariant subspace $\mathcal{J}_\chi^{(\pm)}$, some of the coordinate functions vanish identically. The functions $c_{q_1 q_2}^j(u)$ vanish if $g_{q_2} \in \mathcal{J}_\chi^{(-)}$, $g_{q_1} \notin \mathcal{J}_\chi^{(-)}$ for $q_1 > q_2$ and if $g_{q_1} \in \mathcal{J}_\chi^{(-)}$, $g_{q_2} \notin \mathcal{J}_\chi$ for $q_1 < q_2$. The d functions behave more symmetric. Namely, $d_{q_1 q_2}^j(u)$ vanishes if $g_{q_1} \in \mathcal{J}_\chi^{(+)}$, $g_{q_2} \notin \mathcal{J}_\chi$ and if $g_{q_1} \notin \mathcal{J}_\chi^{(-)}$, $g_{q_2} \in \mathcal{J}_\chi^{(-)}$ for $q_1 > q_2$, and analogously for $q_1 < q_2$. Considered as functions of j, c functions have first order zeros at the points mentioned, whereas d functions have zeros of order $\frac{1}{2}$ (that become of order 1 if they coincide).

Now consider the case that an invariant space $\mathcal{J}_\chi^{(+)}$ exists, let $g_{q_{1,2}}$ both be in $\mathcal{J}_\chi^{(+)}$, and $j < 0$. The normalizing factors N_q appearing in the coordinate functions for the discrete series [see (23)] are already incorporated in the analytic expression for $d_{q_1 q_2}^j(u)$ and we have

$$d_{q_1 q_2}^j(u) = (-1)^{q_1 - q_2}\, c_{q_1 q_2}^{j, \text{discr}}(u). \tag{50}$$

Though it is never necessary to use the d functions, their characteristic symmetries (46), (47), (48), and the simplified appearance of the discrete series, make it sometimes very comfortable to use them. We use them also in order to facilitate the comparison of our results with those in the literature.

4. INVARIANT TRILINEAR FORMS

We consider functions $g_i(\varphi_i) \in \mathcal{D}_{\tau_i}$, $i = 1, 2$, and define the convolution integral operator ($\text{Re}\gamma_k > -1$)

$$K(g_1 \circ g_2)(\varphi_3)$$

$$= \int_0^{2\pi} d\varphi_1 \int_0^{2\pi} d\varphi_2 K(\chi_3, \varphi_3 | \chi_1, \varphi_1; \chi_2, \varphi_2)$$

$$\times g_1(\varphi_1) g_2(\varphi_2) \tag{51a}$$

$$= N(\chi_3 | \chi_1; \chi_2) \int_0^{2\pi} d\varphi_1 \int_0^{2\pi} d\varphi_2$$

$$\times \prod_{\substack{i,j,k \\ \text{cycl.}}} \left(\left|2\sin\frac{\varphi_i - \varphi_j}{2}\right|^{\gamma_k}\right.$$

$$\times \exp[2i\sigma(\varphi_i - \varphi_j)\eta_k]\right) g_1(\varphi_1) g_2(\varphi_2). \tag{51b}$$

The form of the convolution kernel K given in (51b) satisfies the covariance constraints

$$K(g_1 \circ g_2) \in D_{\tau_3},$$
$$T_{v,t}^{\chi_3} K(g_1 \circ g_2) = K(T_{v,t}^{\chi_1} g_1 \circ T_{v,t}^{\chi_2} g_2) \tag{52}$$

provided only that

$$\tau_3 + \eta_1 - \eta_2 = \text{integer} \tag{53}$$

and

$$\gamma_1 = +j_1 - j_2 + j_3 - \tfrac{1}{2},$$
$$\gamma_2 = -j_1 + j_2 + j_3 - \tfrac{1}{2}, \tag{54}$$
$$\gamma_3 = -j_1 - j_2 - j_3 - \tfrac{1}{2}.$$

We require the integrand in (51) to be periodic. This leads to the additional constraints

$$\tau_1 + \eta_3 - \eta_2 = \text{integer},$$
$$\tau_2 + \eta_1 - \eta_3 = \text{integer}. \tag{55}$$

From (53) and (55) follows

$$\tau_1 + \tau_2 - \tau_3 = \text{integer} \tag{56}$$

so that τ_3 is uniquely fixed by τ_1 and τ_2.

If (56) is satisfied, the three constraints (53) and (55) become dependent and one parameter, say η_3, remains free. We set

$$\eta_2 = \tau_1 + \eta_3, \quad \eta_1 = -\tau_2 + \eta_3. \tag{57}$$

Thus it turns out that there is no covariant kernel K of the form (51) if (56) is not fulfilled, but an infinity of them if (56) is fulfilled. However, it will be shown that in the latter case the kernel K can be represented as a linear combination of two fixed kernels K_1 and K_2. This is in agreement with the known fact that the Clebsch–Gordan coefficients for SU(1,1) exhibit a twofold degeneracy.[5,6] It remains to be true in the case of the universal covering group.

The integral (51b) converges absolutely if $\text{Re}\gamma_k > -1$, for all $k = 1, 2, 3$. From (54) it can be seen that this convergence domain contains a strip around the three principal series $\text{Re}j_k = 0$. Thus one defines the operator K first for this strip and then by analytic continuation in the whole space $\mathbb{C}_3$ of the variables j_1, j_2, j_3. As in the case of the intertwining operator, the operator K is singular on a subset of points of $\mathbb{C}_3$. This singular set

is partly connected with the discrete series. A complete discussion of the singular set is not attempted here. We will have to show still that $K(g_1 \circ g_2)(\varphi_3)$, $g_1 \in D_{\tau_1}$, $g_2 \in D_{\tau_2}$ is infinitely differentiable if we are not at a singular point. For $g_3 \in D_{-\tau_3}$

$$\int_0^{2\pi} g_3(\varphi) K(g_1 \circ g_2)(\varphi) d\varphi \tag{58}$$

is then a trilinear invariant form, if g_3 transform as χ_3^d.

Clebsch–Gordan coefficients are defined as the trilinear invariant forms for three elements of the canonical basis. We set

$$g_1(\varphi_1) = g_{q_1}(\varphi_1), \quad g_2(\varphi_2) = g_{q_2}(\varphi_2) \tag{59}$$

and define the "C coefficient":

$$K(g_{q_1} \circ g_{q_2})(\varphi) = (2\pi)^{-1} g_{q_1 + q_2}(\varphi)$$
$$\times C(\chi_3, q_3 | \chi_1, q_1; \chi_2, q_2)_{n_3} \tag{60}$$

for $q_3 = q_1 + q_2$, and

$$C(\chi_3, q_3 | \chi_1, q_1; \chi_2, q_2)_{n_3} = 0 \quad \text{if} \quad q_3 \neq q_1 + q_2. \tag{61}$$

The constraint (56) is automatically taken into account by (61) this way.

In order to express the C coefficient explicitly, it is useful to assume Whipple's notation.[10] Namely

$$F_p(l; m,n) = [\Gamma(\alpha_{ghj}) \Gamma(\beta_{ml}) \Gamma(\beta_{nl})]^{-1}$$
$$\times {}_3F_2(\alpha_{gmn}, \alpha_{hmn}, \alpha_{jmn}; \beta_{ml}, \beta_{nl}; 1), \tag{62}$$

$$F_n(l; m,n) = [\Gamma(\alpha_{lmn}) \Gamma(\beta_{lm}) \Gamma(\beta_{ln})]^{-1}$$
$$\times {}_3F_2(\alpha_{lhj}, \alpha_{lgj}, \alpha_{lgh}; \beta_{lm}, \beta_{ln}; 1) \tag{63}$$

with the labels (g, h, j, l, m, n) denoting any permutation of $(0, 1, 2, 3, 4, 5)$. The α and β coefficients are given in Table I. $F_n(l; m, n)$ and $F_p(l; m, n)$ go into each other under the simultaneous replacements

$$j_k \longleftrightarrow -j_k, \quad q_k \longleftrightarrow -q_k. \tag{64}$$

All $F_p(l; m, n)$ $[F_n(l; m, n)]$ with fixed l represent different series with overlapping convergence domains for the same analytic function of j_k, q_k, $k = 1, 2, 3$ which we denote therefore $F_p(l)$ $[F_n(l)]$. For more details of the relations between all the $F_p(l)$ and $F_n(l)$ see Ref. 10. The $F_p(l; m, n)$ and $F_n(l; m, n)$ are entire analytic functions of the variables α and β and consequently of the

TABLE I. Table of α, β. The coefficients α_{lmn} are totally symmetric whereas $\beta_{mn} = 2 - \beta_{nm}$.

$\alpha_{012} = \tfrac{1}{2} - j_2 + q_2$	$\alpha_{024} = \tfrac{1}{2} - j_1 - j_2 - j_3$	$\alpha_{123} = \tfrac{1}{2} + j_3 + q_3$	$\alpha_{145} = \tfrac{1}{2} + j_1 + j_2 - j_3$
$\alpha_{013} = \tfrac{1}{2} + j_1 - j_2 + j_3$	$\alpha_{025} = \tfrac{1}{2} - j_1 - q_1$	$\alpha_{124} = \tfrac{1}{2} - j_3 + q_3$	$\alpha_{234} = \tfrac{1}{2} - j_1 + q_1$
$\alpha_{014} = \tfrac{1}{2} + j_1 - j_2 - j_3$	$\alpha_{034} = \tfrac{1}{2} - j_2 - q_2$	$\alpha_{125} = \tfrac{1}{2} + j_2 + q_2$	$\alpha_{235} = \tfrac{1}{2} - j_1 + j_2 + j_3$
$\alpha_{015} = \tfrac{1}{2} + j_1 - q_1$	$\alpha_{035} = \tfrac{1}{2} + j_3 - q_3$	$\alpha_{134} = \tfrac{1}{2} + j_1 + q_1$	$\alpha_{245} = \tfrac{1}{2} - j_1 + j_2 - j_3$
$\alpha_{023} = \tfrac{1}{2} - j_1 - j_2 + j_3$	$\alpha_{045} = \tfrac{1}{2} - j_3 - q_3$	$\alpha_{135} = \tfrac{1}{2} + j_1 + j_2 + j_3$	$\alpha_{345} = \tfrac{1}{2} + j_2 - q_2$
$\beta_{01} = 1 - j_1 - j_2 - q_3$	$\beta_{05} = 1 - 2j_2$	$\beta_{15} = 1 + j_1 - j_2 + q_3$	$\beta_{34} = 1 + 2j_3$
$\beta_{02} = 1 + j_1 - j_2 - q_3$	$\beta_{12} = 1 + 2j_1$	$\beta_{23} = 1 - j_1 - j_3 + q_2$	$\beta_{35} = 1 - j_2 + j_3 + q_1$
$\beta_{03} = 1 - j_2 - j_3 - q_1$	$\beta_{13} = 1 + j_1 - j_3 + q_2$	$\beta_{24} = 1 - j_1 + j_3 + q_2$	$\beta_{45} = 1 - j_2 - j_3 + q_1$
$\beta_{04} = 1 - j_2 + j_3 - q_1$	$\beta_{14} = 1 + j_1 + j_3 + q_2$	$\beta_{25} = 1 - j_1 - j_2 + q_3$	

j_k. This can be proved by means of a Barnes's type integral representation.[11]

After a straightforward computation we obtain the C coefficient in the form

$$C(\chi_3, q_1 + q_2 | \chi_1, q_1; \chi_2, q_2)_{\eta_3} = (2\pi)^2 N$$

$$\times \exp(-i\pi(\tau_1 + \tau_2 - \eta_3 - \alpha_{14})/2)$$

$$\times [(\exp(-i\pi\alpha_{045}) - \exp(-i\pi(\beta_{52} + 2\eta_3))) A F_p(1)$$

$$+ (\exp(-i\pi\alpha_{015}) - \exp(-i\pi(\alpha_{045} + \alpha_{345} + 2\eta_3))) B F_p(4)]$$

(65)

with

$$A = \frac{\pi}{\sin\pi\beta_{41}} \frac{\Gamma(\alpha_{023})\Gamma(\alpha_{024})\Gamma(\alpha_{235})}{\Gamma(\alpha_{012})\Gamma(\alpha_{123})} ,$$

(66a)

$$B = \frac{\pi}{\sin\pi\beta_{14}} \frac{\Gamma(\alpha_{013})\Gamma(\alpha_{023})\Gamma(\alpha_{235})}{\Gamma(\alpha_{034})\Gamma(\alpha_{234})} .$$

(66b)

We see that the C coefficient (65) is a linear combination (depending on η_3) of two terms being proportional to $F_p(1)$ or $F_p(4)$. Under the simultaneous replacements

$$j_1 \rightarrow -j_3, \quad j_2 \rightarrow +j_2, \quad q_1 \rightarrow q_3, \quad q_2 \rightarrow -q_2$$

(67)

the subscript of α and β are submitted to the permutation

$$(0, 1, 2, 3, 4, 5) \rightarrow (0, 4, 3, 2, 1, 5).$$

(68)

This amounts to replacing

$$F_p(1) \rightarrow F_p(4), \quad A \rightarrow B.$$

(69)

By a particular choice of η_3 we can either retain only the $F_p(1)$ or the $F_p(4)$ term. Namely, if

$$\eta_{3,1} = -\tfrac{1}{2}\alpha_{245} + \tau_2,$$

(70a)

then the $F_p(1)$ term survives. The convolution operator (51b) for this choice (70a) is denoted K_1 and its C coefficient C_1:

$$C_1(\chi_3; q_1 + q_2 | \chi_1, q_1; \chi_2, q_2) = (2\pi)^3 N(-1)^{q_1 - \tau_1}$$

$$\times \frac{\Gamma(\alpha_{023})\Gamma(\alpha_{024})\Gamma(\alpha_{235})}{\Gamma(\alpha_{012})\Gamma(\alpha_{123})} F_p(1) .$$

(71a)

Similarly we may choose

$$\eta_{3,2} = -\tfrac{1}{2}\alpha_{135}$$

(70b)

and have only the $F_p(4)$ term

$$C_2(\chi_3, q_1 + q_2 | \chi_1, q_1; \chi_2, q_2) = (2\pi)^3 N(-1)^{q_3 - \tau_1 - \tau_2}$$

$$\times \frac{\Gamma(\alpha_{013})\Gamma(\alpha_{023})\Gamma(\alpha_{235})}{\Gamma(\alpha_{034})\Gamma(\alpha_{234})} F_p(4).$$

(71b)

From (68) and (69) we obtain

$$C_1(-j_1, \tau_1, q_1 | -j_3, \tau_3, q_3; j_2, -\tau_2, -q_2)$$

$$= (-1)^{\tau_1 + \tau_2 - \tau_3} C_2(j_3, \tau_3, q_3 | j_1, \tau_1, q_1; j_2, \tau_2, q_2).$$

(72)

Both coefficients C_1, C_2 are polynomially bounded in q_1 and q_2. Infinite differentiability of $K(g_1 \circ g_2)(\varphi_3)$ for $g_1 \in D_{\tau_1}$, $g_2 \in D_{\tau_2}$ can be deduced from this fact.

5. THE DUAL KERNELS AND THE PRINCIPAL SERIES CONTRIBUTION TO THE COMPLETENESS RELATION

Apart from the kernels K_1, K_2, or K we introduce the dual kernel K^d by

$$K^d(g_3)(\varphi_1, \varphi_2) = \int_0^{2\pi} d\varphi_3$$

$$\times K^d(\chi_1, \varphi_1; \chi_2, \varphi_2 | \chi_3, \varphi_3) g_3(\varphi_3).$$

(73)

We postulate covariance

$$K^d(g_3) \in D_{\tau_1} \times D_{\tau_2},$$

(74a)

$$(T_{v,\xi}^{\chi_1} \times T_{v,\xi}^{\chi_2}) K^d(g_3) = K^d(T_{v,\xi}^{\chi_3} g_3).$$

(74b)

It is easily verified that a kernel K^d satisfying these constraints can be obtained from the kernel K (51) by replacing

$$\chi_k \rightarrow \chi_k^d, \quad \eta_k \rightarrow -\eta_k.$$

(75)

Infinite differentiability of $K^d(g_3)(\varphi_1, \varphi_2)$ follows from the polynomial boundedness of the C coefficients in q_k.

We note that for $g_1 \in D_{-\tau_1}$ and $g_2 \in D_{-\tau_2}$ we have a trilinear invariant form

$$\int_0^{2\pi} d\varphi_1 \int_0^{2\pi} d\varphi_2 g_1(\varphi_1) g_2(\varphi_2) K_d(g_3)(\varphi_1, \varphi_2)$$

(76)

when $g_1 (g_2)$ transforms as χ_1^d (χ_2^d). We set

$$\overline{g_1(\varphi_1)} = g_{q_1}(\varphi_1) \in D_{\tau_1},$$

$$\overline{g_2(\varphi_2)} = g_{q_2}(\varphi_2) \in D_{\tau_2},$$

(77)

$$\overline{g_3(\varphi_3)} = g_{q_3}(\varphi_3) \in D_{\tau_3},$$

and denote the value of (76) for these functions by

$$C^d(\chi_1, q_1; \chi_2, q_2 | \chi_3, q_3)_{\eta_3}.$$

(78)

Comparison with (60) yields

$$C^d(\chi_1, q_1; \chi_2, q_2 | \chi_3, q_3)_{\eta_3} = C(\chi_3^d, -q_3 | \chi_1^d, -q_3; \chi_2^d, -q_2)_{-\eta_3}$$

(79)

provided the normalization constant satisfies

$$N^d(\chi_1, \chi_2 | \chi_3) = N(\chi_3^d | \chi_1^d, \chi_2^d).$$

(80)

Thus we have immediately

$$C^d(\chi_1, q_1; \chi_2, q_2 | \chi_3, q_1 + q_2)_{\eta_3}$$

$$= (2\pi)^2 N^d \exp(i\pi(\tau_1 + \tau_2 - \eta_3 + \alpha_{023}/2))$$

$$\times [(\exp(-i\pi\alpha_{123}) - \exp(-i\pi(\beta_{25} - 2\eta_3))) G F_n(1)$$

$$+ (\exp(-i\pi\alpha_{234}) - \exp(-i\pi(\alpha_{123} + \alpha_{012} - 2\eta_3))) H F_n(4)]$$

(81)

with

$$G = \frac{\pi}{\sin\pi\beta_{14}} \frac{\Gamma(\alpha_{145})\Gamma(\alpha_{135})\Gamma(\alpha_{014})}{\Gamma(\alpha_{345})\Gamma(\alpha_{045})} ,$$

(82a)

$$H = \frac{\pi}{\sin\pi\beta_{41}} \frac{\Gamma(\alpha_{245})\Gamma(\alpha_{145})\Gamma(\alpha_{014})}{\Gamma(\alpha_{125})\Gamma(\alpha_{015})} .$$

(82b)

If we fix η_3 such that the kernels K_1 and K_2 result [see (70a) and (70b)], then the two substitutions (75) are incompatible. The phases of C_1^d and C_2^d must therefore be computed explicitly. We choose η_3 so that only the $F_n(1)$ respectively the $F_n(4)$ term survives. From

$$\eta_{3,1} = +\tfrac{1}{2}\alpha_{013} + \tau_2 \tag{83a}$$

we obtain

$$C_1^d = (2\pi)^3 N^d (-1)^{q_1 - \tau_1}$$

$$\times \frac{\Gamma(\alpha_{145})\Gamma(\alpha_{135})\Gamma(\alpha_{014})}{\Gamma(\alpha_{345})\Gamma(\alpha_{045})} F_n(1). \tag{84a}$$

Alternatively

$$\eta_{3,2} = +\tfrac{1}{2}\alpha_{024} \tag{83b}$$

yields

$$C_2^d = (2\pi)^3 N^d (-1)^{q_3 - \tau_1 - \tau_2}$$

$$\times \frac{\Gamma(\alpha_{245})\Gamma(\alpha_{145})\Gamma(\alpha_{014})}{\Gamma(\alpha_{125})\Gamma(\alpha_{015})} F_n(4). \tag{84b}$$

Our aim is the derivation of a completeness relation of the following kind

$$\int d\mu\,(\chi_3)_{PS} \int_0^{2\pi} d\varphi_3$$

$$[\Delta_1(\chi_1,\chi_2,\chi_3) K_2^d(\chi_1,\varphi_1;\ \chi_2,\varphi_2\,|\,\chi_3,\varphi_3\,)$$

$$\times K_1(\chi_3,\varphi_3\,|\,\chi_1,\varphi_1';\ \chi_2,\varphi_2')$$

$$+ \Delta_2(\chi_1,\chi_2,\chi_3) K_1^d(\chi_1,\varphi_1;\ \chi_2,\varphi_2\,|\,\chi_3,\varphi_3)$$

$$\times K_2(\chi_3,\varphi_3\,|\,\chi_1,\varphi_1';\ \chi_2,\varphi_2')]$$

$$+\ \text{discrete series contribution}$$

$$= \sum_{k_1,2 = -\infty}^{+\infty} \exp(2\pi i(\tau_1 k_1 + \tau_2 k_2))$$

$$\times \delta(\varphi_1 - \varphi_1' - 2\pi k_1)\delta(\varphi_2 - \varphi_2' - 2\pi k_2). \tag{85}$$

We assume that $\chi_{1,2}$ belongs to the principal series first. Other representations will be reached by analytic continuation later. We have to derive Δ_1, Δ_2 and the measure $d\mu\,(\chi_3)_{PS}$ on the principal series in the j_3 plane.

In addition we have to show that combinations $K_1^d K_1$ and $K_2^d K_2$ that could be allowed by covariance do not occur in (85).

We take matrix elements of (85) and obtain

$$\delta_{q_1 + q_2,\,q_1' + q_2'} \int d\mu\,(\chi_3)_{PS}$$

$$[\Delta_1 C_2^d(\chi_1,q_1;\ \chi_2,q_2\,|\,\chi_3,q_1 + q_2)$$

$$\times C_1(\chi_3,q_1' + q_2'\,|\,\chi_1,q_1';\ \chi_2,q_2')$$

$$+ \Delta_2 C_1^d(\chi_1,q_1;\ \chi_2,q_2\,|\,\chi_3,q_1 + q_2)$$

$$\times C_2(\chi_3,q_1' + q_2'\,|\,\chi_1,q_1';\ \chi_2,q_2')]$$

$$+\ \text{discrete series contribution}$$

$$= (2\pi)^3 \delta_{q_1 q_1'} \delta_{q_2 q_2'}. \tag{86}$$

This relation is obtained as special case of a more general formula.

Following Ferretti and Verde[5] and Wang,[6] we use "coordinate functions of the second kind":
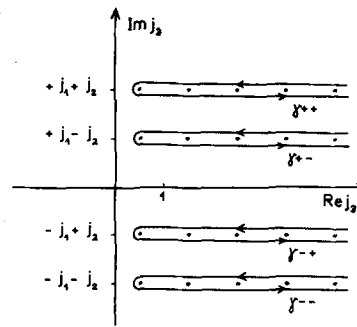
FIG. The contours $\gamma_{\pm,\pm}$.

$$e_{qq'}^j(z) = \frac{\pi}{\sin\pi(-2j)} \{j\}^{1/2} [\tfrac{1}{2}(z+1)]^{-(q+q')/2}$$

$$\times [\tfrac{1}{2}(z-1)]^{(q+q')/2 - j - 1/2}$$

$$\times [\Gamma(\tfrac{1}{2} - j - q')\Gamma(\tfrac{1}{2} - j + q)\Gamma(2j+1)]^{-1}$$

$$\times\, {}_2F_1(\tfrac{1}{2} + j - q,\ \tfrac{1}{2} + j - q';\ 2j+1;\ 2/(1-z)), \tag{87}$$

$$e_{qq'}^j(z) = (-1)^{q-q'} e_{q'q}^j(z) = (-1)^{q-q'} e_{-q,-q'}^j(z)$$

$$\times [\arg(z+1) = \arg(z-1) = 0]. \tag{88}$$

Here $\{j\}^{1/2}$ is the analytic function of j described in Sec. 3 [Eq. (43)]. One considers then the product

$$e_{q_1 q_1'}^{j_1}(z)\, e_{q_2 q_2'}^{j_2}(z)$$

and expands it into a sum over

$$e_{q_1 + q_2,\,q_1' + q_2'}^{j_3(n)}(z),\quad j_3(n) = \tfrac{1}{2} + j_1 + j_2 + n$$

by means of the Burchnall—Chaundy formula.[12] Since $e_{q_3,q_3'}^{j_3}(z)$ falls off exponentially in the right half j_3 plane [see the Appendix, Eq. (A15)] we can transform this sum into a contour integral. Then we add up four such integrals by means of

$$d_{qq'}^j(z) = e_{qq'}^j(z) + e_{qq'}^{-j}(z) \tag{89}$$

to obtain $(z > 1)$, $q_1 \geqslant q_1'$, $q_2 \geqslant q_2'$

$$d_{q_1 q_1'}^{j_1}(z)\, d_{q_2 q_2'}^{j_2}(z) = \frac{1}{2i} \int_{\sum \gamma_{\pm,\pm}} dj_3\, 2j_3$$

$$\times [\tan\pi(j_3 + \tau_3) + \tan\pi(j_3 - \tau_3)] E(j_3) e_{q_1 + q_2,\,q_1' + q_2'}^{j_3}(z) \tag{90}$$

with

$$E(j_3) = \{j_1\}^{1/2}\{j_2\}^{1/2}\{j_3\}^{1/2} \frac{\pi}{\sin\pi(-2j_2)}$$

$$\times \left(\frac{\Gamma(\alpha_{135})\Gamma(\alpha_{145})\Gamma(\alpha_{235})\Gamma(\alpha_{245})}{\Gamma(\alpha_{012})\Gamma(\alpha_{034}')} F_p(0) F_n'(5) \right.$$

$$\left. - \frac{\Gamma(\alpha_{013})\Gamma(\alpha_{014})\Gamma(\alpha_{023})\Gamma(\alpha_{024})}{\Gamma(\alpha_{125})\Gamma(\alpha_{345}')} F_p(5) F_n'(0) \right). \tag{91}$$

The contours $\gamma_{\pm,\pm}$ are depicted in Fig. 2. Primed (unprimed) terms depend only on primed (unprimed) variables q_k.

A Sommerfeld—Watson deformation of the contour $\sum_{\pm}\gamma_{\pm,\pm}$ which is justified in the Appendix, leads to

$$\int_{\sum_{\pm}\gamma_{\pm,\pm}} \to -\int_{-i\infty}^{+i\infty} + \sum_{\text{Res}}.$$

Use is here made of the fact that the functions $F_p(1)$,

$F_n(1)$ are free of singularities in j_3. Because the imaginary axis and all the integrands except the factor $e^{j_3}_{q_3 q'_3}$ are symmetric under $j_3 \to -j_3$, we can replace $e^{j_3}_{q_3 q'_3}$ by $\frac{1}{2} d^{j_3}_{q_3 q'_3}$ in (90). Together this yields ($z \geqslant 1$)

$$d^{j_1}_{q_1 q'_1}(z) d^{j_1}_{q_2 q'_2}(z) = -\frac{1}{2i} \int^{+i\infty}_{-i\infty} dj_3 j_3$$

$$+ [\tan\pi(j_3 + \tau_3) + \tan\pi(j_3 - \tau_3)] E(j_3) d^{j_3}_{q_3 q'_3}(z)$$

$$+ \sum_{0 \leqslant n \leqslant [q'_3 - 1/2]} 2j_3(n) E(j_3(n)) d^{j_3(n)}_{q_3 q'_3}(z) \qquad (92)$$

with

$$j_3(n) = q'_3 - n - \tfrac{1}{2}. \qquad (93)$$

The sum of residues stems from the poles of the function $e^{j_3}_{q_3 q'_3}(z)$.

The sum of the residue terms has been given for the case

$$q_3 \geqslant q'_3 \geqslant \tfrac{1}{2}. \qquad (94)$$

If

$$-\tfrac{1}{2} \geqslant q_3 \geqslant q'_3, \qquad (95)$$

then we have a sum over n from 0 to

$$[-q_3 - \tfrac{1}{2}] \quad \text{and} \quad j_3(n) = -q_3 - n - \tfrac{1}{2}. \qquad (96)$$

If neither (94) nor (95) is fulfilled (still for $q_3 \geqslant q'_3$), then the sum is void. The sum of residues represents the contributions of the discrete series.

Introducing the real parameter ρ by

$$j_3 = i\rho, \qquad (97)$$

we can rewrite the measure on the imaginary axis

$$d\mu(\chi_3)_{\text{PS}}$$

$$= -\frac{1}{2i} j_3 [\tan\pi(j_3 + \tau_3) + \tan\pi(j_3 - \tau_3)] dj_3$$

$$= \frac{\rho \sinh 2\pi\rho}{\cosh 2\pi\rho + \cos 2\pi\tau_3} d\rho. \qquad (98)$$

Thus the measure $d\mu(\chi_3)_{\text{PS}}$ is positive.

The relation (92) is usually derived first under the assumption

$$q_1 \geqslant q'_1, \quad q_2 \geqslant q'_2. \qquad (99)$$

If we go to the limit $z = 1$ [absolute convergence of the integral in (92) will be proven for $z = 1$ in the Appendix], the factor

$$\delta_{q_1 + q_2, q'_1 + q'_2} \text{ [on the rhs of (92)] allows only}$$

$$q_1 = q'_1, \quad q_2 = q'_2.$$

In order to obtain orthogonality of the C coefficients, say, for $q_1 \neq q'_1$, we have to abandon (99), say, by considering

$$q_1 \geqslant q'_1, \quad q_2 \leqslant q'_2. \qquad (100)$$

However, the derivation of (92) uses the coordinate functions of the second kind, whose hypergometric factor is symmetric in q, q'. Since we want to maintain the connection

$$q_1 + q_2 = q_3, \quad q'_1 + q'_2 = q'_3, \qquad (101)$$

it can be seen that the coefficients in the Burchnall—Chaundy expansion remain unchanged. It turns out that $E(j_3)$ is independent of the order and that (92) is valid in general.

With the help of the three-term relations for $F_p(1)$ and $F_n(1)$,[10] we replace $F_p(0)$, $F_p(5)$, $F'_n(5)$, $F'_n(0)$ in $E(j_3)$ by $F_n(1)$, $F_n(4)$, $F'_p(1)$, $F'_p(4)$ and get

$$E(j_3) = \{j_1\}^{1/2} \{j_2\}^{1/2} \{j_3\}^{1/2}$$

$$\times [V F_n(1) F'_p(4) + W F_n(4) F'_p(1)] \qquad (102)$$

with

$$V = \frac{+\pi^3}{\sin\pi\beta'_{14} \sin\pi\alpha_{145} \sin\pi\alpha_{235}}$$

$$\times \frac{\Gamma(\alpha_{013})\Gamma(\alpha_{135})}{\Gamma(\alpha'_{045})\Gamma(\alpha'_{034})\Gamma(\alpha_{234})\Gamma(\alpha'_{345})}, \qquad (103a)$$

$$W = \frac{-\pi^3}{\sin\pi\beta_{14} \sin\pi\alpha_{145} \sin\pi\alpha_{235}}$$

$$\times \frac{\Gamma(\alpha_{024})\Gamma(\alpha_{245})}{\Gamma(\alpha_{012})\Gamma(\alpha'_{015})\Gamma(\alpha_{123})\Gamma(\alpha_{125})}. \qquad (103b)$$

Comparison of (102) with (86) then yields

$$\Delta_1(\chi_1, \chi_2, \chi_3) = -\Delta_2(\chi_1, \chi_2, \chi_3)$$

$$= (2\pi)^{-3}(NN^d)^{-1} \frac{\pi}{\sin\pi(j_1 + j_3 + \tau_2)}. \qquad (104)$$

We have still to determine the discrete series contributions.

6. THE DISCRETE SERIES CONTRIBUTION TO THE COMPLETENESS RELATION

The contribution of the discrete series to (85), (92) involves the discrete measure $2j_3$ concentrated at the positions (93) if $q_3 \geqslant 0$ or (96) if $q_3 \leqslant 0$. This measure depends on q_3. The introduction of the covariant kernels is therefore more complicated than in the case of the principal series.

The easiest way to deal with the discrete series makes use of kernels K_1, K_3, K^d_1, K^d_3 (which we are going to define) instead of K_1, K_2, K^d_1, K^d_2 in (85). In fact, the choice (70a,b) is not unique for retaining a simple Whipple factor $F_p(l)$ or $F_n(l)$ in the C coefficients. One can isolate any $F_p(l)$ or $F_n(l)$ by means of three-term relations[10] if one wants to. But the function $\eta_3(j_3)$ one has to choose is not always elementary. Other elementary choices are, e.g.,

$$K_3: \eta_3 = -\tfrac{1}{2}\alpha_{014} - \tau_1, \qquad (105)$$

$$C_3 = (2\pi)^3 N(-1)^{q_2 - \tau_2} \frac{\Gamma(\alpha_{013})\Gamma(\alpha_{023})\Gamma(\alpha_{024})}{\Gamma(\alpha_{025})\Gamma(\alpha_{035})} F_p(5), \qquad (106)$$

$$C^d_3 = (2\pi)^3 N^d(-1)^{q_2 - \tau_2} \frac{\Gamma(\alpha_{135})\Gamma(\alpha_{145})\Gamma(\alpha_{245})}{\Gamma(\alpha_{124})\Gamma(\alpha_{134})} F_n(5), \qquad (107)$$

$$K_4: \eta_3 = +\tfrac{1}{2}\alpha_{014} - \tau_1, \qquad (108)$$

$$C_4 = (2\pi)^3 N(-1)^{q_2 - \tau_2} \frac{\Gamma(\alpha_{013})\Gamma(\alpha_{023})\Gamma(\alpha_{024})}{\Gamma(\alpha_{123})\Gamma(\alpha_{234})} F_n(0), \qquad (109)$$

$$C^d_4 = (2\pi)^3 N^d(-1)^{q_2 - \tau_2} \frac{\Gamma(\alpha_{135})\Gamma(\alpha_{145})\Gamma(\alpha_{245})}{\Gamma(\alpha_{015})\Gamma(\alpha_{045})} F_p(0). \qquad (110)$$

The function $E(j_3)$ [(91)] can be written in terms of C_1, C_3, C_1^d, C_3^d easily. The principal series contribution in (85) then looks like

$$\int d\mu(\chi_3)_{\text{PS}} \int_0^{2\pi} d\varphi_3 [\Delta_3(\chi_1,\chi_2) K_3^d(\chi_1, \varphi_1; \chi_2, \varphi_2 | \chi_3, \varphi_3)$$

$$\times K_1(\chi_3, \varphi_3 | \chi_1, \varphi_1'; \chi_2, \varphi_2') + \Delta_4(\chi_1, \chi_2)$$

$$\times K_1^d(\chi_1, \varphi_1; \chi_2, \varphi_2 | \chi_3, \varphi_3) K_3(\chi_3, \varphi_3 | \chi_1, \varphi_1'; \chi_2, \varphi_2')] \quad (111)$$

with the weight factor

$$\Delta_3(\chi_1, \chi_2) = -\Delta_4(\chi_1, \chi_2)$$

$$= (2\pi)^{-3} (NN^d)^{-1} \frac{\pi}{\sin\pi(j_1 - j_2 + \tau_1 + \tau_2)} \quad (112)$$

that can be extracted from the integral since it is independent of j_3 [contrary to $\Delta_{1,2}$ in (104)]. Correspondingly $E(j_3)$ is proportional to

$$\Delta_3(\chi_1, \chi_2) C_3^d C_1' + \Delta_4(\chi_1, \chi_2) C_1^d C_3'. \quad (113)$$

If $q_3 \geq 0$, then for the discrete series terms $\Gamma(\alpha_{035})^{-1} = 0$ and consequently $C_3 = 0$. Thus, only the first term in (113) survives. We introduce a new kernal $\hat{K}_3^d$ as the convolution of K_3^d with an intertwining operator, precisely

$$\hat{C}_3^d(\chi_1, q_1; \chi_2, q_2 | \chi_3^c, q_3) = \frac{\Gamma(\alpha_{124})}{\Gamma(\alpha_{123})} C_3^d(\chi_1, q_1; \chi_2, q_2 | \chi_3, q_3). \quad (114)$$

The contribution of positive q_3 to the discrete series can then be presented in the form

$$+ \Delta_3(\chi_1, \chi_2) \sum_{j_3} 2j_3 \int_0^{2\pi} d\varphi_3 \int_0^{2\pi} d\varphi_3'$$

$$\times \hat{K}_3^d(\chi_1, \varphi_1; \chi_2, \varphi_2 | \chi_3^c, \varphi_3)$$

$$\times S(\chi_3^c, \varphi_3 | \chi_3, \varphi_3') K_1(\chi_3, \varphi_3' | \chi_1, \varphi_1'; \chi_2, \varphi_2'). \quad (115)$$

The intertwining kernel S is obtained from (26) by a special normalization and by setting

$$\tau_3 \stackrel{\scriptscriptstyle\wedge}{=} j_3 + \tfrac{1}{2} \mod 1. \quad (116)$$

This gives explicitly

$$S(\chi_3^c, \varphi_3 | \chi_3, \varphi_3') = [\Gamma(1 + 2j_3)/2\pi]$$

$$\times [-2i \sin\tfrac{1}{2}(\varphi_3 - \varphi_3' + i0)]^{-2j_3 - 1}. \quad (117)$$

In fact, the sum over j_3 in (115) runs over all positive j_3 so that

$$\tau_1 + \tau_2 \stackrel{\scriptscriptstyle\wedge}{=} j_3 + \tfrac{1}{2} \mod 1 \quad (118)$$

is fulfilled.

Negative q_3 can be treated analogously. In this case (96) holds and for the discrete series terms $\Gamma(\alpha_{123})^{-1} = 0$ implying $C_1 = 0$. Consequently, the second term in (113) survives. We define

$$\hat{C}_1^d(\chi_1, q_1; \chi_2, q_2 | \chi_3^c, q_3) = [\Gamma(\alpha_{045})/\Gamma(\alpha_{035})]$$

$$\times C_1^d(\chi_1, q_1; \chi_2, q_2 | \chi_3, q_3). \quad (119)$$

As contribution of negative q_3 to the discrete series we obtain this way

$$+ \Delta_4(\chi_1, \chi_2) \sum_{j_3} 2j_3 \int_0^{2\pi} d\varphi_3 \int_0^{2\pi} d\varphi_3'$$

$$\times \hat{K}_1^d(\chi_1, \varphi_1; \chi_2, \varphi_2 | \chi_3^c, \varphi_3) \overline{S(\chi_3^c, \varphi_3 | \chi_3, \varphi_3')}$$

$$\times K_3(\chi_3, \varphi_3' | \chi_1, \varphi_1'; \chi_2, \varphi_2'). \quad (120)$$

The intertwining kernel S [its complex conjugate enters (120)!] is the same as in (117). The sum runs over all positive j_3 obeying

$$-\tau_1 - \tau_2 \stackrel{\scriptscriptstyle\wedge}{=} j_3 + \tfrac{1}{2} \mod 1. \quad (121)$$

The completeness relation says that the sum of (111), (115), and (120) gives the delta functions

$$\sum_{k_1 = -\infty}^{+\infty} \exp(2\pi i \tau_1 k_1) \delta(\varphi_1 - \varphi_1' - 2\pi k_1)$$

$$\times \sum_{k_2 = -\infty}^{+\infty} \exp(2\pi i \tau_2 k_2) \delta(\varphi_2 - \varphi_2' - 2\pi k_2). \quad (122)$$

APPENDIX: ASYMPTOTIC BEHAVIOR AND CONVERGENCE OF INTEGRALS

We make use of the asymptotic expression

$$\Gamma(a + j)/\Gamma(b + j) \sim j^{a-b}(1 + O(1/j)),$$

$$|\arg j| \leq \pi - \epsilon \quad (A1)$$

following from Stirling's formula. The line $\arg j = \pi$, can be included as follows. We have

$$\frac{\Gamma(a + j)}{\Gamma(b + j)} = \frac{\sin\pi(b + j)}{\sin\pi(a + j)} \frac{\Gamma(1 - b - j)}{\Gamma(1 - a - j)} \quad (A2)$$

and from (A1)

$$\Gamma(1 - b - j)/\Gamma(1 - a - j) \sim j^{a-b}(1 + O(1/j)),$$

$$|\arg(-j)| \leq \pi - \epsilon. \quad (A3)$$

If we keep away from the poles of $\Gamma(a + j)$ by a fixed distance ϵ, then from (A1), (A2), (A3)

$$|\Gamma(a + j)/\Gamma(b + j)| \leq M(\epsilon)(1 + |j|)^{\text{Re}(a-b)} \quad (A4)$$

everywhere. In fact for a complex number $x = \xi + i\eta$ we have

$$|\sin\pi x|^2 = \sin^2\pi\xi + \sinh^2\pi\eta,$$

$$|\sin\pi x| \geq \max(|\sin\pi\xi|, |\sinh\pi\eta|). \quad (A5)$$

In the case of hypergeometric functions $_3F_2$ at the argument one similar reasoning leads to valuable estimates. We start with the following.

Lemma: The analytic functions of x defined from the hypergeometric series

$$_3F_2(a, b, c; d, e + x; 1), \quad (A6a)$$

$$_3F_2(a, b, c + x; d + x, e + x; 1) \quad (A6b)$$

by analytic continuation in x, have asymptotic expansions for $|x| \to \infty$, valid at least for $|\arg(x - x_0)| \leq \pi/2$ (x_0 arbitrary fixed), that are termwise equal with the hypergeometric series themselves.

We have been able to prove (A6a) for $|\arg(x - x_0)| \leq \pi/2 - \epsilon$. A general proof does not seem to be known yet (see Ref. 13). We rely on the fact that it is plausible. It can be verified easily in the special cases where the hypergeometric series can be summed.

In the left half planes the hypergeometric functions have poles at the same positions as the functions

$$\Gamma(e+x)\Gamma(e+d+x-a-b-c) \tag{A7a}$$

respectively

$$\Gamma(d+x)\Gamma(e+x)\Gamma(e+d+x-a-b-c). \tag{A7b}$$

Assuming validity of the lemma allows us to extend the asymptotic expansion into the left half plane up to $|\arg x| \leqslant \pi - \epsilon$ or to the whole left half plane except parallel strips around the poles. Moreover, an estimate is possible in the same sense as the estimate (A4). These results are achieved by inverting the sign of x in (A6a) by a three-term relation.

If we take $E(j_3)$ in the form (91), we meet the functions $F_p(0)$, $F_p(5)$, $F_n(0)$, $F_n(5)$ which contain hypergeometric factors of the type

$$_3F_2(a,b,c;\, d+x,\, e-x;\, 1) \tag{A8}$$

that posses poles in both half planes. Again a three-term relation allows us to express the function (A8) by two functions (A6b) or two functions (A6b) with the sign of x inverted. Since we are interested in an estimate of $E(j_3)$ for $|\arg j_3| \leqslant \pi/2$ we need the functions (A6b) themselves.

In fact, from the definition we have

$$_3F_2(a,b,c;\, d_0+x,\, e_0-x;\, 1)$$
$$= \Gamma(s)\Gamma(d_0+x)\Gamma(e_0-x)F_p(0,4,5), \tag{A9}$$

where the parameters α_{ghj}, β_{mn} are connected with the parameters $a,b,c,d=d_0+x$, $e=e_0-x$, $s=d+e-a-b-c$, as in Ref. 10, Table 4.1. Then we express $F_p(0)$ by $F_p(5)$ and $F_n(4)$ using the three-term relation

$$F_p(0) = \Gamma(1-e+a)\Gamma(1-e+b)\Gamma(1-e+c)$$

$$\times \left(\frac{\sin\pi e}{\pi\Gamma(s)}\, F_n(4) + \frac{1}{\Gamma(a)\Gamma(b)\Gamma(c)}\, F_p(5)\right) \tag{A10}$$

and realize $F_n(4)$ and $F_p(5)$ by

$$F_n(4) = F_n(4;\, 0,1), \quad F_p(5) = F_p(5;\, 0,1), \tag{A11}$$

to which the lemma applies. It follows
$$_3F_2(a,b,c;\, d_0+x,\, e_0-x;\, 1) \sim 1 + O(1/x)$$

$$+ (\pi/\sin\pi e)[\Gamma(s)/\Gamma(a)\Gamma(b)\Gamma(c)]$$

$$\times x^{-1+2(a+b+c)-e_0-d_0}(1+O(1/x)), \quad |x| \to \infty, \quad |\arg x| \leqslant \pi/2. \tag{A12}$$

The second term can obviously be neglected if $|\mathrm{Im}\,x| \to \infty$. Similar situations arise if the functions (A6a), (A6b) are studied in the left half plane: for $\mathrm{Im}\,x = \mathrm{const}$ there results a power behavior which vanishes if $|\mathrm{Im}\,x| \to \infty$, leaving the asymptotic expansion in a series of inverse powers as expected from the lemma.

Treating the first term in (91) with this techniques we obtain (including the factors in front of the bracket)

$$\sim \{j_1\}^{1/2}\{j_2\}^{1/2}\, \frac{\pi}{\sin\pi(-2j_2)}\, (\mp\, ij_3)^{-2+2(q_1'-q_1)+q_3-q_3'}$$

$$\times (1+O(1/j_3)), \quad |\mathrm{Im}\,j_3| \to \infty, \quad |\arg j_3| \begin{array}{c} \geqslant \epsilon \\ \leqslant \pi - \epsilon \end{array}. \tag{A13}$$

For $\mathrm{Im}\,j_3 = \mathrm{const}$ we get some power behavior if we keep a distance ϵ away from the poles. The second term in (91) gives exactly the same result as (A13) and both terms cancel. This could be expected since $E(j_3)$ is independent of the order of $q_1 \gtrless q_1'$, $q_2 \gtrless q_2'$. Consequently, there must be cancellations up to the result

$$E(j_3) \sim M(j_1, j_2)(\mp\, ij_3)^{-3-|q_1-q_1'|-|q_2-q_2'|}\, (1+O(1/j_3)),$$

$$|\mathrm{Im}\,j_3| \to \infty, \quad |j_3| \to \infty. \tag{A14}$$

Finally we take into account the exponential decrease of the second kind function

$$[\tan\pi(j_3+\tau_3) + \tan\pi(j_3-\tau_3)]\{j_3\}^{1/2}\, e_{qq'}^{j_3}(z)$$

$$\sim -\frac{2}{\pi\sinh\eta}^{1/2} j_3^{q-q'-1/2}\exp(-\eta j_3), \tag{A15}$$

$$\eta = \mathrm{arc}\,\cosh z > 0, \quad |\arg j_3| \leqslant \pi - \epsilon.$$

This justifies the replacement of the Burchnall—Chaundy series by the contour integral over $\sum_{\pm}\gamma\pm$, $\pm$ and the Sommerfeld—Watson deformation of this contour. For $\mathrm{Re}\,j_3 = 0$ we have uniformly in z

$$\left|d_{qq'}^{j_3}(z)\right| \leqslant 1, \quad z \geqslant 1 \tag{A16}$$

following from the unitarity of the principal series. This justifies the limit $z \to 1$.

[1]V. Bargmann, Ann. Math. **48**, 568 (1947).

[2]L. Pukánszky, Trans. Am. Math. Soc. **100**, 116 (1961).

[3]I.M. Gel'fand, M.I. Graev, and N.Ya. Vilenkin, *Generalized Functions*, Vol. 5 (Academic, New York and London, 1966), Chap. VII.

[4]W. Rühl, *The Lorentz Group and Harmonic Analysis* (Benjamin, New York, 1970), Chaps. 5, 6.

[5]I. Ferretti and M. Verde, Nuovo Cimento **55**, 110 (1968); M. Verde, Accademia Nazionale dei Lincei, Atti del Covegno Mendeleeviano, Torino and Rome, September 1969, edited by V. Bona, Torino, 1971.

[6]K.-H. Wang, J. Math. Phys. **11**, 2077 (1970).

[7]M. Andrews and J. Gunson, J. Math. Phys. **5**, 1391 (1964).

[8]W.J. Holman and L.Biedenharn, Ann. Phys. (N.Y.) **39**, 1 (1966); **47**, 205 (1968).

[9]L. Schwartz, *Théorie des distributions* (Hermann, Paris 1966), Chap. VI, § 8.

[10]L.J. Slater, *Hypergeometric Functions* (Cambridge U.P., Cambridge, 1966), Chap. IV.

[11]Ref. 10, Eq. (4.2.2.1).

[12]J.L. Burchnall and T.W. Chaundy, Proc. London Math. Soc. **50**, 56 (1944).

[13]Y.L. Luke, *The Special Functions and Their Approximations*, Mathematics in Science and Engineering, Vol. 53 (Academic, New York and London, 1969), esp. Chap. VII; U.J. Knottnerus, *Approximation Formulae for Generalized Hypergeometric Functions for Large Values of the Parameters* (Wolters, Groningen, 1960).

Lagrange multipliers and gravitational theory

J. L. Safko and F. Elston

University of South Carolina, Columbia, South Carolina 29208
(Received 23 March 1976)

The Lagrange multiplier version of the Palatini variational principle is extended to nonlinear Lagrangians, where it is shown in the case of the quadratic Lagrangians, as expected, that this version of the Palatini approach is equivalent to the Hilbert variational method. The (nonvanishing) Lagrange multipliers for the quadratic Lagrangians are then explicitly obtained in covariant form. It is then pointed out how the Lagrange multiplier approach in the language of the (3+1)-formalism developed by Arnowitt, Deser, and Misner permits the recasting of the equations of motion for quadratic and general higher-order Lagrangians into the ADM canonical formalism. In general without the Lagrange multiplier approach, the higher order ADM problem could not be solved. This is done explicitly for the simplest quadratic Lagrangian ($g^{1/2}R^2$) as an example.

I. INTRODUCTION

Since the earliest days of general relativity it was known that the field equations of general relativity could be obtained from a Hilbert variational principle using the Einstein Lagrangian, $g^{1/2}R$. The Hilbert variational approach assures the connection is Riemannian and treats the Lagrangian as a function of $g_{\mu\nu}$ and its derivatives. An alternative method, the Palatini approach was proposed almost as early as the Hilbert method.[1] In the Palatini approach the metric and the affine connection are varied independently and the Riemann relationship between the metric and the connection is deduced as a consequence. The Palatini approach is known to be equivalent to the Hilbert approach for the Einstein Lagrangian $g^{1/2}R$.

However, the Palatini approach gives different results than the Hilbert method when we apply it to higher order Lagrangians. In particular the Lagrangians $g^{1/2}R^2$, $g^{1/2}R_{\mu\nu}R^{\mu\nu}$, and $g^{1/2}R_{\alpha\beta\gamma\delta}R^{\alpha\beta\gamma\delta}$ have been investigated by several authors using the Palatini method.[2-4] In all cases the field equations are different than those obtained from the Hilbert approach. This difficulty has been handled either by simply choosing the quantities $g_{\mu\nu}$ to be the metric, as a special case, after the variation is performed, or by simply leaving the two sets of equations obtained by separate variations of the affine connection and the $g_{\mu\nu}$ as independent sets of equations in a general affine space.

Lanczos[5] and later Ray[6] have shown that the Palatini procedure for $g^{1/2}R$ can be understood as a Lagrange multiplier approach where it turns out that the Lagrange multiplier itself is zero. We believe that the Palatini treatment of Stephenson,[2] Higgs,[3] and Roxburgh[4] of the quadratic Lagrangian of a Riemannian space are unnecessary and incorrect. In Sec. II we will show that extending the idea of Lagrange multipliers to nonlinear actions leads to the expected results with the additional information that the Lagrange multipliers are no longer identically zero. We will examine the quadratic Lagrangians as particular examples. In Sec. III we will apply a Lagrange multiplier approach to the Arnowitt, Deser, and Misner (ADM) 3 + 1 canonical formalism. We show that a Lagrange multiplier approach enables us to formally solve the ADM approach for nonlinear actions, which was not possible in the original ADM method. We consider the Lagrangian $g^{1/2}R^2$ as a particular example.

The Lagrange multipliers used in Sec. III are of a completely different nature than those used in Sec. II. The constraint of Sec. II must be pre-imposed in Sec. III since all the affine symbols involving time as an index must be explicitly expressed in terms of Christoffel symbols in order to produce a (3 + 1)-formalism. In the (3 + 1)-formalism we use the Lagrange multiplier approach to decouple the variations within the intrinsic 3-geometry (expressed in terms of ${}^3g_{ij}$ and ${}^3\Gamma^i_{jk}$) from the variation involving the extrinsic curvature (expressed in terms of K_{ij}, N_i, N). This procedure is "natural" to the (3 + 1)-formalism because it emphasizes the break up between the intrinsic geometry of the 3-space and the extrinsic curvature of the 3-space with respect to the enveloping 4-space.

Before proceeding with these investigations we need to discuss why we wish to consider nonlinear actions. There are, in general, 14 independent algebraic invariants of the Riemann tensor, any one of a combination of which could be used as the Lagrangian of a variational principle that would yield a "completely geometrodynamical" theory,[1] one expressed entirely in terms of the structure of the curvature of space—time. Classical general relativity results, of course, from the variation of one of the simplest of these invariants, the Ricci scalar R. In recent years, the next most complicated invariants, the three independent quadratic invariants R^2, $R_{\mu\nu}R^{\mu\nu}$, and $R_{\alpha\beta\gamma\delta}R^{\alpha\beta\gamma\delta}$ have attracted renewed attention in several respects. They have been treated in connection with the question of stability against gravitational collapse[7] and also mentioned in connection with the renormalization difficulties of attempts to quantize the gravitational field.[8,9] In view of this interest, we first clarify some confusion in the current literature regarding the method of obtaining field equations from the quadratic Lagrangians by showing (in Sec. II) that the alternative methods of "Hilbert" and "Palatini" variational procedures can be made to yield *equivalent* results if the Palatini method is correctly interpreted as a Lagrange multiplier technique as indicated by Lanczos[5] and Ray.[6]

Since one of the major lines of approach to the quan-

tization of general relativity is the "canonical Hamiltonian formalism" developed by ADM, it should prove desirable also to be able to express the field equation of the quadratic invariants in this formalism. We show in Sec. III that a "Lagrange multiplier" procedure here provides a very natural way to re-express the equations of motion into a first-order canonical formalism, and exhibit the explicit form of the field equations in this formalism for the simplest quadratic Lagrangian, $g^{1/2}R^2$.

The idea of using Lagrange multipliers is reminiscent of the attempts of Lanczos to develop a canonical formalism for nonlinear Lagrangians.[10] His approach, however, was quite different from the ADM method. The multipliers he introduced were of a different nature than those we introduce in the section on the ADM approach.

The final section (IV) contains a discussion of the general applicability of the Lagrange multiplier method in the 4-formalism and discusses the questions of interpretation of the nonvanishing of the Lagrange-multiplier constraint forces, with some suggestions for future work.

II. LAGRANGE MULTIPLIERS—4-FORMALISM FOR NONLINEAR ACTIONS

A. General comments on Lagrange multipliers

Many papers have been written applying the normal Palatini approach to nonlinear actions[2-4] such as $g^{1/2}R^2$, $g^{1/2}R^{\alpha\beta}R_{\alpha\beta}$, etc. If one followed Ray's suggestion,[6] then the correct way to approach such problems would seem to be via Lagrange multipliers. With such an approach it is obvious that the field equations so obtained for $g_{\mu\nu}$ must be equivalent to the normal Hilbert variation. However, we will have the additional information contained in the Lagrange multipliers.

Suppose we wish to examine some action

$$A = \int g^{1/2} L \, d^2x, \tag{1}$$

where the integral is over all 4-space and L is the Lagrangian of the system. The normal Einstein approach is to calculate the set of equations.

$$\delta A / \delta g_{\mu\nu} = 0. \tag{2}$$

We instead consider

$$A = \int g^{1/2} L \, d^4x + \int g^{1/2} L_c \, d^4x, \tag{3}$$

where

$$L_c = \Lambda_\alpha{}^{\beta\gamma}[\Gamma^\alpha{}_{\beta\gamma} - \tfrac{1}{2}g^{\alpha\rho}(g_{\rho\beta,\gamma} + g_{\rho\gamma,\beta} - g_{\beta\gamma,\rho})] \tag{4}$$

with $\Lambda_\alpha{}^{\beta\gamma}$ Lagrange multipliers.

We now solve the set of equations

$$\delta A / \delta g_{\mu\nu} = 0, \tag{4a}$$

$$\delta A / \delta \Gamma^\alpha{}_{\beta\gamma} = 0, \tag{4b}$$

$$\delta A / \delta \Lambda_\alpha{}^{\beta\gamma} = 0, \tag{4c}$$

where the variations are taken treating $g_{\mu\nu}$, $\Gamma^\alpha{}_{\beta\gamma}$, and $\Lambda_\alpha{}^{\beta\gamma}$ as independent quantities. Equation (4c) simply recovers the constraint. Solving Eqs. (4c) and (4b) and substituting into Eq. (4a) must, in principle, produce the same result as Eq. (2). However, we now have the additional information contained in the Lagrange multiplier given by Eq. (4b).

To calculate δL_c, we need the following identity for affine spaces in which the connection implied by Eq. (4) is not known:

$$\delta[\tfrac{1}{2}g^{\alpha\rho}(g_{\rho\beta,\gamma} + g_{\rho\gamma,\beta} - g_{\beta\gamma,\rho})]$$
$$= \tfrac{1}{2}(g_{\rho\beta;\gamma} + g_{\rho\gamma;\beta} - g_{\beta\gamma;\rho})\delta g^{\alpha\rho} + \tfrac{1}{2}g^{\alpha\rho}[(\delta g_{\rho\beta})_{;\gamma} + (\delta g_{\rho\gamma})_{;\beta}$$
$$- (\delta g_{\beta\gamma})_{;\rho}]. \tag{5}$$

This identity can easily be established for affine spaces with symmetric connection by an extension of Palatini's original argument to affine spaces. As long as the connection is symmetric in its covariant indices, by a theorem due to Weyl,[11] there exists a locally geodesic coordinate system in which ordinary and covariant differentiation are equivalent. In this coordinate system $g_{\beta\gamma,\rho} = g_{\beta\gamma;\rho}$ and, since the variation operations and ordinary differentiation commute, we see that $\delta(g_{\beta\gamma;\rho}) = (\delta g_{\beta\gamma})_{;\rho}$. This variation is a tensor so the explicitly covariant form of the preceeding must hold in all coordinate systems. This establishes Eq. (5) for all affine spaces with symmetric connections. Once the constraint of a Riemannian space is imposed by Eq. (4c), the terms involving the covariant derivatives of $g_{\mu\nu}$ vanish and Eq. (5) reduces to the well-known result for a Riemann space.

Noting also that terms such as $\int L (\delta g) \, d^4x$ vanish when Eq. (4c) is applied, we can now discuss the effects of the constraint upon Eqs. (4a) and (4b).

For (4b) we obtain

TABLE I. Lagrange multipliers and constraint effects on quadratic actions.

Case	L	$\dfrac{\delta L}{\delta g_{\mu\nu}}\bigg|_{\Gamma^\alpha{}_{\beta\gamma}=\text{const}}$	$\Lambda_\alpha{}^{\beta\gamma}$	$B^{\mu\nu}$
A	R^2	$-2RR^{\mu\nu} + \tfrac{1}{2}g^{\mu\nu}R^2$	$-2g^{\beta\gamma}R_{;\alpha}$	$-R_{;}{}^{\mu\nu} - R_{;}{}^{\nu\mu} + g^{\mu\nu}R_{;}{}^\sigma{}_\sigma$
B	$R_{\mu\nu}R^{\mu\nu}$	$\begin{cases} -2R^\mu{}_\sigma R^{\sigma\nu} \\ +\tfrac{1}{2}g^{\mu\nu}R_{\sigma\tau}R^{\sigma\tau} \end{cases}$	$-2R^{\beta\gamma}{}_{;\alpha}$	$\begin{cases} -R^{\mu\sigma,\nu}{}_{;\sigma} - R^{\nu\sigma}{}_{;}{}^\mu{}_\sigma \\ +R^{\mu\nu,\sigma}{}_{;\sigma} \end{cases}$
C	$R_{\alpha\beta\gamma\delta}R^{\alpha\beta\gamma\delta}$	$\begin{cases} -2R^\mu{}_{\sigma\tau\rho}R^{\nu\sigma\tau\rho} \\ +\tfrac{1}{2}g^{\mu\nu}R_{\alpha\beta\gamma\delta}R^{\alpha\beta\gamma\delta} \end{cases}$	$\begin{cases} -4R_\alpha{}^{\beta\gamma\tau}{}_{;\tau\sigma} \\ -4R_\alpha{}^{\gamma\beta\tau}{}_{;\tau} \end{cases}$	$\begin{cases} +2R^{\mu\sigma\nu\tau}{}_{;\tau\sigma} \\ +2R^{\nu\sigma\mu\tau}{}_{;\tau\sigma} \end{cases}$

$$\left(\frac{\delta L}{\delta\Gamma^\alpha{}_{\beta\gamma}}+\Lambda_\alpha{}^{\beta\gamma}\right)\delta\Gamma^\alpha{}_{\beta\gamma}=0 \tag{6}$$

while Eq. (4a) becomes

$$[\delta L/\delta g_{\mu\nu}+\tfrac{1}{2}(\Lambda^{\mu\nu\sigma}{}_{;\sigma}+\Lambda^{\nu\mu\sigma}{}_{;\sigma}-\Lambda^{\sigma\mu\nu}{}_{;\sigma})]\delta g_{\mu\nu}=0. \tag{7}$$

The variation in Eqs. (6) and (7) are taken considering $g_{\mu\nu}$ and $\Gamma^\alpha{}_{\beta\gamma}$ as independent quantities.

It is now obvious that Eq. (6) can always be solved explicitly for the Lagrange multipliers $\Lambda_\alpha{}^{\beta\gamma}$ and substituted into Eq. (7). As Ray pointed out,[5] the $\Lambda_\alpha{}^{\beta\gamma}\equiv 0$ if $L=R$; however, this is not true in general. The particular examples of the quadratic scalar invariants of a Riemannian space are shown in Table I. We denote the second quantity in Eq. (7): i.e.,

$$B^{\mu\nu}=\tfrac{1}{2}(\Lambda^{\mu\nu\sigma}{}_{;\sigma}+\Lambda^{\nu\mu\sigma}{}_{;\sigma}-\Lambda^{\sigma\mu\nu}{}_{;\sigma}). \tag{8}$$

$B^{\mu\nu}$ is obviously symmetric in μ and ν, since we took $\Gamma^\alpha{}_{\beta\gamma}$ symmetric in β and γ. It is clear that from adding columns 3 and 5 of Table I that the results obtained are equivalent to those of a straightforward Hilbert variation such as found in DeWitt.[9]

B. Discussion of "normal" Palatini results for quadratic Lagrangians

Stephenson[2] has considered the quadratic Lagrangians of Table I with a "normal" Palatini variation. This is equivalent to our results with $\Lambda_\alpha{}^{\beta\gamma}\equiv 0$, and hence $B^{\mu\nu}=0$. The three entries in our table are labeled A, B, C in agreement with Stephenson's cases A, B, C. In a later paper Higgs[3] shows that the only solutions to Stephenson's cases A and B are the Einstein equations with or without cosmological constant. Roxburgh[4] has arrived at essentially the same conclusion for a combination of cases A and B for symmetric affine connections.

As Buchdahl[12] has pointed out there are objections to the ambiguities and "unphysical" aspects of applying the Palatini method as it is normally given. In the light of of our analysis we can add another "physical" objection. If we follow the usual Palatini procedure, then the set of Riemannian spaces that satisfy Stephenson's equations is much more limited than those that satisfy the Hilbert variational procedure. Stephenson's, Higgs', and Roxburgh's solutions correspond to choosing $\Lambda_\alpha{}^{\beta\gamma}=0$. From Table I we see that for case A, $\Lambda_\alpha{}^{\beta\gamma}=0$ corresponds to $R_{;\alpha}=0$ which leads to Higgs' conclusion[3] if we assume the normal relation between the connection and the metric. Likewise, for case B, $\Lambda_\alpha{}^{\beta\gamma}=0$ corresponds to $R^{\beta\gamma}{}_{;\alpha}=0$ which again leads to Higgs' conclusion for Riemannian spaces.

For the Lagrange multiplier approach to reduce to the results of Stephenson $et\ al.$ it is not necessary that $\Lambda_\alpha{}^{\beta\gamma}=0$ but only that

$$B^{\mu\nu}=0. \tag{9}$$

This condition on the sum of the covariant derivatives is less stringent than that imposed by the previous requirement.

In classical physics Lagrange multipliers are related to forces of constraint. It would thus seem appropriate to call the 40 quantities, $\Lambda_\alpha{}^{\beta\gamma}$, generalized forces of constraint that confine physical 4-space to be those symmetric affine spaces that are Riemannian. That is, there is a choice for the 40 Lagrange multipliers such that the space is Riemannian. If one imagines the space composed of all possible symmetric affine spaces, then the set of all Riemannian spaces is a hypersurface of this space and the Lagrange multipliers constrain us to this hypersurface. It is possible that greater understanding of the nature of the spaces defined by the higher order invariants can be obtained by examining these "constraint forces."

III. LAGRANGE MULTIPLIERS AND THE ADM FORMALISM

A. The connection between the Lagrange multiplier approach and the ADM formalism

In Sec. II we saw that the Lagrange multiplier approach, although equivalent to the Hilbert variational method, does simplify calculations in the 4-formalism approach to general relativity. In our attempt to extend the $3+1$ canonical formalism of ADM to higher order invariants, the Lagrange multipliers will play a much more essential role. As ADM point out, two essential aspects of the canonical form are[13]

(1) that the field equations are first order in the time derivatives, and

(2) that time has been singled out so that the theory has been recast into $(3+1)$-dimensional form.

The Riemann tensor is second-order in the time derivatives, so all of the 14 independent scalar invariants constructible from it are also second-order in time. For the particular case of the Einstein Lagrangian, $g^{1/2}R$, it is possible to neatly sidestep this difficulty and satisfy condition (1) by eliminating a total time derivative (as well as a divergence term for further simplification) and use the equivalent Lagrangian[13]

$$L_{\text{ADM}}=\dot g_{ij}\pi^{ij}-NR^0-N_iR^i, \tag{10}$$

which is already first order (although of second degree) in the time derivatives.

Alternatively, we could use the full Einstein Lagrangian expressed in terms of g_{ij} and its conjugate π^{ij} and do a "Palatini variation" as ADM indicate.[13] When they take this approach, ADM have built in the constraint that the 3-geometry is Riemannian. This artifice is successful in achieving canonical form only because of a peculiarity of the Lagrangian used. Even the full Einstein Lagrangian $^4g^{1/2}R$ is only linear in $\dot\pi^{ij}$, and so the Lagrange equations for g_{ij} and π^{ij} as independent variables happen to be in canonical form. All higher order Lagrangians constructed by the ADM "Palatini approach" will be nonlinear in $\dot\pi^{ij}$; hence, some other approach is needed to achieve a canonical Hamiltonian form for the equations of motion. In order to use this approach we will need the extrinsic curvature K_{ij}.

In the remainder of this section we will show how to apply the Lagrange multiplier approach introduced in Sec. II to the ADM formalism in order to enable us to

put the higher-order invariants into canonical form. This is not trivial for the higher-order invariants. After some general discussion in this part of Sec. III we will examine the Einstein Lagrangian in subsection B and finally look at the simplest quadratic Lagrangian, $g^{1/2}R^2$, in subsection C.

Since all the Riemann invariants are second-order in time, they would in 3+1 formalism have the general functional form

$$\mathcal{L} = \mathcal{L}(^3g_{ij}, K_{ij}, \dot{K}_{ij}, N, N_i),$$

where the extrinsic curvature tensor K_{ij} is related to the intrinsic 3-geometry $^3g_{ij}$ of the spacelike hypersurface by its definition:

$$K_{ij} = (1/2N)(N_{i|j} + N_{j|i} - {}^3\dot{g}_{ij}) \tag{11}$$

No Riemann scalar invariant will contain any higher-order time derivatives than $\dot{K}_{ij}$. It would seem that for the variation process we can give up the definition of K_{ij} in terms of $\dot{g}_{ij}$, and treat the K_{ij} as independent co-ordinates. These coordinates appear in the Lagrangian only in first order in their first time derivative. We will impose, as a nonholonomic constraint, the definition of K_{ij} in terms of $\dot{g}_{ij}$. This constraint is of the form:

$$\delta \int \mathcal{L}_{\text{constraint}} {}^4g^{1/2}d^4x$$
$$= \delta \int \lambda^{ij}[K_{ij} - (1/2N)(N_{i|j} + N_{j|i} - {}^3\dot{g}_{ij})]N \, {}^3g^{1/2}d^3x \, dt \tag{12}$$

It should be noted that the "constraint" on K_{ij} here is not related to the Riemannian nature of the over-all 4-space. The extrinsic curvature tensor K_{ij} is an entirely spacelike 3×3 tensor which is completely specified once we have given the intrinsic 3-geometry of a spacelike slice, imposed the initial-value equations on it, and chosen a coordinate system. Thus it is already "constrained" in form. Because of this, we cannot simply treat it as a "free" variable as a direct usual "Palatini"-type approach would do; if we wish to treat it as a free variable in the variation, then we must use the Lagrange multiplier method to express the "constraints" on its form separately. The Lagrange multipliers λ^{ij} must then be nonvanishing always since the "constraint" (i.e., the restriction on the form of K_{ij}) is always real. We will find that the λ^{ij} are nonzero even for the case of the Einstein Lagrangian $^4g^{1/2}\,{}^4R$. This is simply a consequence of using the Lagrange multiplier formalism to "decouple" the variation with respect to K_{ij} from that variations of $^3g_{ij}$, N, and N_i, and in itself implies nothing about the Riemannian nature of the over-all 4-space.

Essentially, in order to write the Lagrangian in (3+1)-form at all, the constraint of K_{ij} in terms of N, $N_{ij}\,{}^3g_{ij}$ has to be "pre-imposed" (i.e., 4-space Riemannian as is spacelike 3-space), and we do not solve for those constraints in this formalism. They are implicit in the initial-value equations obtained by varying N and N_i, since it is N, N_i, and $^3g_{ij}$ that determine the imbedding once a coordinate system is chosen.

We now construct a new variational principle.

$$\delta \int (\mathcal{L}_R + \mathcal{L}_c)N^3g^{1/2}d^3x \, dt = 0 \tag{13}$$

where $\mathcal{L}_c$ is Eq. (12) above and $\mathcal{L}_R$ is one of the 14 independent scalar invariants of a Riemann space. The quantity

$$\mathcal{L}_R = \mathcal{L}_R(^3g_{ij}, K_{ij}, \dot{K}_{ij}, N, N_i)$$

is now to be considered a function of two independent sets of variables $^3g_{ij}$ and K_{ij} and their first order time derivatives. Thus, the total Lagrangian has been rendered first order in time so that we can proceed to the canonical Hamiltonian formalism in the usual way. The 3-space itself must remain a three-dimensional Riemannian manifold with a metric signature $(+++)$. One can choose the 3-space initially Riemannian by either varying the $^3g_{ij}$ directly by a Hilbert variational principle or by a Lagrange multiplier technique similar to Sec. II.

As we will discuss in subsection B, if the scalar Lagrangian $\mathcal{L}_R$ is chosen to be R, it is possible to eliminate $\dot{K}_{ij}$ and hence the Lagrangian multiplier is trivial to find. In all other cases (we examine R^2 in detail in subsection B) the invariant is sufficiently complicated that any attempt to eliminate $\dot{K}_{ij}$ is doomed to fail. Thus it is in general essential to use the Lagrange multiplier approach of Eq. (12) for the higher-order Riemann invariants if one wishes to recast them into a 3+1 or ADM canonical formalism.

It is also interesting to note that the constraint of Eq. (11) is of an interesting mathematical type—a nonholonomic constraint. In fact, it is the simplest such constraint. That is, one variable (K_{ij}) is related linearly to the first time derivative of another corresponding variable $(^3g_{ij})$.

B. The Einstein Lagrangian, $^4g^{1/2}\,{}^4R$

For simplicity we will begin with the ADM equivalent Lagrangian for this case:

$$\mathcal{L}_{\text{ADM}} = N\,{}^3g^{1/2}\,{}^3R - N\,{}^3g^{1/2}[(\text{Tr}K)^2 - \text{Tr}(K^2)],$$

where $\text{Tr}K = K^j{}_j$ and $\text{Tr}(K^2) = K^i{}_jK^j{}_i$. This is first-order and second degree in g_{ij} but is written entirely in terms of $^3g_{ij}$, K_{ij}, and the lapse N and shift N_i. We can treat the K_{ij} as separate coordinates if we add the constraint Lagrangian

$$\mathcal{L}_c = N\,{}^3g^{1/2}\lambda^{ij}K_{ij} - \tfrac{1}{2}\,{}^3g^{1/2}\lambda^{ij}(N_{i|j} + N_{j|i})$$
$$+ \tfrac{1}{2}\,{}^3g^{1/2}\lambda^{ij}\dot{g}_{ij},$$

where the λ^{ij} are the Lagrange multipliers.

The momenta π^{ij} and Π^{ij} conjugate to $^3g_{ij}$ and K_{ij} respectively are trivial to evaluate:

$$\pi^{ij} = \frac{\partial \mathcal{L}_{\text{total}}}{\partial \dot{g}_{ij}} = \tfrac{1}{2}g^{1/2}\lambda^{ij}, \tag{14}$$

$$\Pi^{ij} = \frac{\partial \mathcal{L}_{\text{total}}}{\partial \dot{K}_{ij}} = 0, \tag{15}$$

where we have dropped the superscript 3 on the $g^{1/2}$ and in all future cases where the quantities meant are clear. Equation (14) allows us to eliminate the Lagrange multiplier. We can now form the Hamiltonian density

$$H = \dot{g}_{ij}\pi^{ij} + \dot{K}_{ij}\Pi^{ij} - L_{\text{total}}$$

$$= -Ng^{1/2}R + Ng[(\text{Tr}K)^2 - \text{Tr}(K^2)]$$

$$- 2N\pi^{ij}K_{ij} + (N_{i|j} + N_{j|i})\pi^{ij} + \dot{K}_{ij}\Pi^{ij} \qquad (16)$$

and then the Hamiltonian H by

$$H = \int H \, d^3x.$$

Some of Hamilton's equations recover definitions

$$\dot{g}_{ij} = \frac{\delta H}{\delta \pi^{ij}} = -2NK_{ij} + N_{i|j} + N_{j|i}, \qquad (17)$$

which are the constraint equations, and

$$\dot{K}_{ij} = \frac{\delta H}{\delta \Pi^{ij}} = \dot{K}_{ij}. \qquad (18)$$

The latter Eqs. (18) are trivial in this case only because the equivalent Lagrangian Eq. (10) did not depend upon the $\dot{K}_{ij}$. In general, Eq. (18) will recover the definition of Π^{ij} in terms of $\dot{K}_{ij}$. The equation in $\dot{g}_{ij}$, Eq. (17), will be the same no matter what scalar invariant we take for the Lagrangian since it depends only upon the constraint. The dynamical Hamiltonian equations are

$$\dot{\pi}^{ij} = -\frac{\delta H}{\delta g_{ij}}, \qquad (19)$$

$$\dot{\Pi}^{ij} = -\frac{\delta H}{\delta K_{ij}} = 0, \qquad (20)$$

where the zero in Eq. (20) is again a result of the particular invariant (^4R) that we used.

We now have

$$\dot{\pi}^{ij} = Ng^{1/2}(R^{ij} - \tfrac{1}{2}g^{ij}R) + 2Ng^{1/2}(K^{im}K_m{}^j - K^m{}_mK^{ij}) \quad (21)$$

and

$$\dot{\Pi}^{ij} = 2Ng^{1/2}(K^{ij} - g^{ij}K) + 2N\pi^{ij} = 0. \qquad (22)$$

We can make the identification

$$\pi^{ij} = \pi^{ij}{}_{\text{ADM}} = g^{1/2}(g^{ij}K^m{}_m - K^{ij}). \qquad (23)$$

Since the λ^{ij} are related to the π^{ij} in Eq. (14) we see that we cannot take $\lambda^{ij} = 0$ even for the Einstein invariant if we use a canonical formalism. That is, there is no direct "Palatini" approach unless we use a constraint, either "pre-imposed" or added on by Lagrange multipliers, in the 3 + 1 canonical formalism. Since the Lagrange multiplier is effectively equivalent to the momenta for any action there is no simplification possible in our approach.

In addition to the dynamical equations, general relativity (and other geometrodynamic theories obtained from the higher-order invariants) is unusual in that it also has constraints on the initial value data. This data must be consistent with the initial value data. This data must be consistent with the initial value equations, which are obtained by varying the action integral, Eq. (13) with respect to N and N_i (lapse and shift). We may in a sense consider N and N_i as "Lagrange multipliers" which give us the initial constraints. For the Einstein Lagrangian we obtain

$$\frac{\delta A}{\delta N} = g^{1/2}\{R + [\text{Tr}(K^2) - (\text{Tr}K)^2]\} + 2\pi^{ij}K_{ij} = 0 \qquad (24)$$

and

$$\frac{\delta A}{\delta N_i} = 2\pi^{ij}{}_{|j} = 0, \qquad (25)$$

which are identical with the ADM results since $\pi^{ij} = \pi^{ij}{}_{\text{ADM}}$. These initial value equations are not expressed in terms of the Hamiltonian because they are not dynamical equations. They are only constraints on the initial value data.

Thus, in the 3 + 1 canonical formulation of general relativity the Lagrange multipliers have a different physical interpretation than we found in the 4-formulation. For the 4-formulation the Lagrange multipliers $\Lambda_\alpha{}^{\mu\nu}$ were zero and represented the constraints needed to keep 4-space Riemann while in the canonical approach those Lagrange multipliers relating $\dot{g}_{ij}$ and K_{ij}, i.e., the λ^{ij}, are related to the canonical momentum.

It is interesting to note for our "truncated" Lagrangian L_{ADM} (with constraint terms L_c and $L_{\text{total}} = L_{\text{ADM}} + L_c$) that when we make the identification $\pi^{ij} = \tfrac{1}{2}g^{1/2}\lambda^{ij}$, then the "Hamilton's" equations of motion, Eqs. (21) and (22), are in the fact exactly the same as the Lagrange's equations for g_{ij} and K_{ij} equations, while our "defining" equation for π^{ij}, Eq. (17), is just the constraint equation obtained by varying L_c with respect to λ^{ij} (this last identity holds for any Lagrangian). Equation (18) for $\dot{K}_{ij}$ in this case yields nothing, so that our "Hamiltonian" equations and "Lagrange's" equations for this "truncated" Einstein Lagrangian are in fact identical. It should be remarked that the automatic Riemannian nature of the space—time for the Einstein Lagrangian $^4g^{1/2}\,{}^4R$ obtained in Sec. II can still also be seen in the context of the 3 + 1 formalism. If we throw in Lagrange multiplier constraints for the 3-space Christoffel symbols $^3\Lambda_i{}^{jk}$ ($^3\Gamma^i{}_{jk} - \tfrac{1}{2}g^{im}(g_{mj,k} + g_{mk,j} - g_{jk,m}))Ng^{1/2} = L_c$, then the variational analysis within the 3-space proceeds exactly as in Sec. II for the 4-space and we obtain explicitly $^3\Lambda_i{}^{jk} = 0$. Hence, an arbitrary spacelike slice is automatically Riemannian within the intrinsic 3-space. Further, we note that the equations of motion, Eq. (22), for the time evolution of Π^{ij} implies that $\pi^{ij} = \pi^{ij}_{\text{ADM}}$ and our initial value equations are consistent with those obtained by ADM with the pre-imposed condition. Since the above considerations hold for any arbitrary spacelike slice through the structure of 4-space, we conclude that the over-all 4-space for the Einstein Lagrangian must be automatically Riemannian if the Lagrangian is the Einstein $g^{1/2}R$.

C. The quadratic Lagrangian $g^{1/2}R^2$

The quadratic Lagrangian, $^4g^{1/2}\,{}^4R^2$, represents the simplest quadratic Riemannian invariant used as an action for a possible field theory. Thus, it will provide the simplest example where the Lagrangian multiplier technique is needed. In this Lagrangian we will take the $g_{ij} \equiv (^3g_{ij})$ and the K_{ij} as independent variables and impose the relation between K_{ij} and $\dot{g}_{ij}$ by Lagrange multipliers so that the Lagrangian density is

$$L_{\text{total}} = L_{R^2} + Ng^{1/2}\lambda^{ij}[K_{ij} - (1/2N)(N_{i|j} + N_{j|i} - \dot{g}_{ij})]. \qquad (26)$$

1535   J. Math. Phys., Vol. 17, No. 8, August 1976

J.L. Safko and F. Elston   1535

All geometric quantities without a preceeding superscript are three-dimensional quantities.

When we form $L_{R^2} = ({}^4g^{1/2}\,{}^4R)^2/g^{1/2}$, we must use the full expression for the Einstein Lagrangian ${}^4g^{1/2}\,{}^4R$ rather than the truncated expression L_{ADM} with the divergence and total time derivative removed, since the left-out terms will now contribute nonvanishing cross terms when we form $({}^4g^{1/2}\,{}^4R)^2$. The full expression for the entire Einstein Lagrangian is[1]

$${}^4g^{1/2}\,{}^4R = -g_{ij}\dot{\pi}^{ij}_{\text{ADM}} - Ng^{-1/2}[\text{Tr}(\pi^2_{\text{ADM}}) - \tfrac{1}{2}(\pi_{\text{ADM}})^2] + Ng^{1/2}R$$
$$+ 2N_i(\pi^{ij}_{\text{ADM}})_{|j} - 2[\pi^{ij}_{\text{ADM}}N_j - \tfrac{1}{2}N^i\pi_{\text{ADM}} + N^{|i}g^{1/2}]_{,i},$$

where $\pi^{ij}_{\text{ADM}} = g^{1/2}(g^{ij}K - K^{ij})$ and $\pi_{\text{ADM}} = g_{ij}\pi^{ij}_{\text{ADM}}$. In order to form $g^{1/2}R^2$, we eliminate π^{ij}_{ADM} and $\dot{\pi}^{ij}_{\text{ADM}}$ in favor of K_{ij} and $\dot{K}_{ij}$. If we once again let $K = \text{Tr}K_{ij} = K^i{}_i$, $\text{Tr}(K^2) = K_j{}^mK_m{}^j$, we can express L_{R^2} simply provided we use the notation

$$A = -2g^{1/2}g^{ij}\dot{K}_{ij}, \qquad B = +Ng^{1/2}K^2,$$
$$C = -3Ng^{1/2}\text{Tr}(K^2), \qquad D = +Ng^{1/2}R, \qquad (27)$$
$$E = +2g^{1/2}(N_{i|j} + N_{j|i})K^{ij}, \qquad F = +2g^{1/2}N^iK_{|i},$$
$$G = -2g^{1/2}N^{|i}{}_{|i}, \qquad Q = 1/Ng^{1/2}.$$

Then, using ${}^4g^{1/2}\,{}^4R^2 = ({}^4g^{1/2}\,{}^4R)^2/{}^4g^{1/2}$, we obtain

$$L_{R^2}/Q = A^2 + 2A(B + C + D + E + F + G)$$
$$+ (B + C + D + E + F + G)^2, \qquad (28)$$

where the only terms dependent upon $\dot{K}_{ij}$ are first two since only they contain the term A. We will identify them as $L_{\dot{K}_{ij}}/Q = A^2 + 2A(B + C + D + E + F + G)$. These are the terms which prevent us from proceeding with the canonical form without recourse to Lagrange multipliers.

Using Eq. (12) for the constraint, we obtain again, as we must,

$$\pi^{ij} = \frac{\partial L_{\text{total}}}{\partial \dot{g}_{ij}} = \tfrac{1}{2}g^{1/2}\lambda^{ij}, \qquad (29)$$

but now the momentum conjugate to K_{ij} is not zero but is

$$\Pi^{ij} = \frac{\partial L_{\text{total}}}{\partial \dot{K}_{ij}} = \frac{8}{N}g^{1/2}g^{ij}(g^{rs}\dot{K}_{rs})$$
$$- \frac{4}{N}g^{ij}(B + C + D + E + F + G). \qquad (30)$$

This enables us to express $L_{\dot{K}_{ij}}$ as

$$L_{\dot{K}_{ij}} = \dot{K}_{ij}\Pi^{ij} - (4/N)g^{1/2}(g^{ij}\dot{K}_{ij})^2. \qquad (31)$$

Using Eq. (30) for Π^{ij} enables us to eliminate $g^{ij}\dot{K}_{ij}$ from Eq. (31), producing

$$L_{\dot{K}_{ij}} = \dot{K}_{ij}\Pi^{ij} - [Ng^{-1/2}/16K^2](K_{ij}\Pi^{ij})^2$$
$$- [g^{-1/2}/2K](B + C + D + E + F + G)(K_{ij}\Pi^{ij})$$
$$- [1/Ng^{1/2}](B + C + D + E + F + G)^2 \qquad (32)$$

The last term cancels all the remaining terms in the L_{R^2} of Eq. (28), so that

$$L_{R^2} = \dot{K}_{ij}\Pi^{ij} - [Ng^{-1/2}/16K^2](K_{ij}\Pi^{ij})^2$$
$$- [g^{-1/2}/2K](B + C + D + E + F + G)(K_{ij}\Pi^{ij}). \qquad (33)$$

Once again the initial value equations are well posed and come from the total action

$$A = \int L_{\text{total}}\,dt = \int L_{\text{total}}\,d^3x\,dt$$

so that

$$\frac{\delta A}{\delta N} = \begin{bmatrix} + 2\pi^{ij}K_{ij} - (g^{-1/2}/16K^2)(K_{rs}\Pi^{rs})^2 \\ - (g^{-1/2}/2K)(K_{rs}\Pi^{rs})[g^{1/2}K^2 - 3g^{1/2}\text{Tr}(K^2) + g^{1/2}R] \\ + \{(K_{rs}\Pi^{rs})g^{ij}/K\}_{|i|j} \end{bmatrix}$$
$$= 0 \qquad (34)$$

and

$$\frac{\delta A}{\delta N_i} = \begin{bmatrix} + 2\pi^{ij}{}_{|j} + 2\{(K_{rs}\Pi^{rs})K^{ij}/K\}_{|j} \\ - (K_{rs}\Pi^{rs})[K^{|i}/K] \end{bmatrix} = 0. \qquad (35)$$

The problem is now straightforward. Defining $L_{\text{total}} = L_{R^2} + L_{\text{constraint}}$ and treating g_{ij} and K_{ij} separately, we form the Hamiltonian as $H = \int H\,d^3x$, where the Hamiltonian density H is $H = \dot{g}_{ij}\pi^{ij} + \dot{K}_{ij}\Pi^{ij} - L_{\text{total}}$. There are four different Hamilton's equations. Two recover the definitions of π^{ij} and Π^{ij}: i.e.,

$$\dot{g}_{ij} = \frac{\delta H}{\delta\pi^{ij}} = -2NK_{ij} + (N_{i|j} + N_{j|i}), \qquad (36)$$

which recovers our constraint of K_{ij} in terms of N, N_i, and $\dot{g}_{ij}$, and

$$\dot{K}_{ij} = \frac{\delta H}{\delta\Pi^{ij}} = +\frac{1}{8}\frac{Ng^{-1/2}}{(K)^2}(K_{rs}\Pi^{rs})K_{ij}$$
$$+ \frac{g^{-1/2}}{2(K)}(B + C + D + E + F + G)K_{ij}, \qquad (37)$$

where taking the trace of Eq. (37) and solving for Π^{ij} shows that it indeed simply recovers the definition of Π^{ij} as $\partial L_{\text{total}}/\partial\dot{K}_{ij}$.

The remaining two Hamilton's equations, the dynamical equations describing the time evolution of the 3-space, are

$$\dot{\Pi}^{ij} = -\delta H/\delta K_{ij}$$
$$= +2N\pi^{ij} + \tfrac{1}{8}(Ng^{-1/2}/K^3)(K_{rs}\Pi^{rs})^2g^{ij} - \tfrac{1}{8}(Ng^{-1/2}/K^2)$$
$$\times(K_{rs}\Pi^{rs})\Pi^{ij} + \tfrac{1}{2}(g^{-1/2}/K^2)(K_{rs}\Pi^{rs})g^{ij}$$
$$\times[B + C + D + E + F + G] - \tfrac{1}{2}(g^{-1/2}/K)\Pi^{ij}$$
$$\times[B + C + D + E + F + G]$$
$$- (1/K)\begin{bmatrix} + NKg^{ij} - 3NK^{ij} \\ + (N^{i|j} + N^{j|i}) \end{bmatrix} + \{(K_{rs}\Pi^{rs})N^mg^{ij}/K\}_{|m}, \qquad (38)$$

$$\dot{\pi}^{ij} = -\delta H/\delta g_{ij}$$
$$= +\tfrac{1}{32}(Ng^{-1/2}/K^2)(K_{rs}\Pi^{rs})^2g^{ij}$$
$$+ \tfrac{1}{8}[Ng^{-1/2}(K_{rs}\Pi^{rs})^2/K^3]K^{ij} + [B + C + D + E + F + G]$$
$$\times\begin{bmatrix} (+\tfrac{1}{4}(g^{-1/2}/K)(K_{rs}\Pi^{rs})g^{ij} \\ + \tfrac{1}{2}(g^{-1/2}/K^2)(K_{rs}\Pi^{rs})K^{ij} \end{bmatrix} + (K_{rs}\Pi^{rs})(\tfrac{1}{2}/K)$$
$$\times\begin{bmatrix} -\tfrac{1}{2}NK^2g^{ij} - 2NKK^{ij} \\ + \tfrac{3}{2}N\text{Tr}(K^2)g^{ij} + 6NK^{im}K_m{}^j \\ - N(R^{ij} - \tfrac{1}{2}g^{ij}R) - K^{rs}(N_{r|s} + N_{s|r})g^{ij} \\ - N^mK_{|m}g^{ij} - 2N^mK^{ij}{}_{|m} + N^{|m}{}_{|m}g^{ij} \end{bmatrix}$$

$$+\left\{(K_{rs}\Pi^{rs})N^m K^{ij}/K\right\}_{|m}. \qquad (39)$$

Although these are extremely complicated expressions, it is clear that the formal conversion from 4-notation to (3 + 1)-notation has been completed, and that the same process will work using any order scalar invariant of the Riemann tensor in the action principle.

IV. CONCLUSIONS

We have seen, then, that the classical Lagrange multiplier technique proves of significant value in discussing the field equations of higher-order Lagrangians in two distinct ways:

(1) In the covariant 4-formalism treated in Sec. II, we can extend the "Palatini" procedure to higher-order invariants by employing the Lagrange multiplier technique to constrain the affine connections to be Christoffel symbols of a Riemann space. The equations of motion obtained are, as expected, equivalent to those obtained directly from the Hilbert method; however, we gain the advantage of being able to solve explicitly for the Lagrange multiplier constraint forces, and thus identify those terms in the equations of motion which arise directly from the constraint to Riemann space.

(2) In the 3 + 1 formalism treated in Sec. III, we can use the Lagrange multiplier technique to "decouple" the variations of the intrinsic 3-geometry of a 3-space from the variations of the extrinsic curvature K_{ij} describing the embedding of the 3-space within the enveloping four-dimensional space. This allows us to recast the equations of motion into a first-order (in time) canonical Hamiltonian formalism (as developed by ADM) suitable for attempts at conventional quantization techniques.

The purposes achieved by the Lagrange multiplier technique in each of these formalisms (Covariant 4-formalism and ADM (3 + 1)-formalism) in the context of the higher-order invariants are complementary in the following sense: In the 4-formalism, we can solve easily and directly for the constraint forces that keep the space—time Riemannian, but we obtain a second order in time Lagrangian formalism that is unsuitable for quantization purposes, whereas in the (3 + 1)-formalism the Lagrange multipliers are not related to the constraints to a Riemann space, which must be pre-imposed in order to achieve (3 + 1)-form for the Lagrangians, but do allow us to recast the equations of motion into a first-order canonical Hamiltonian formalism. Further, although the present paper concerns only the quadratic

invariants of the Riemann tensor, it is obvious that the techniques of Secs. II and III will apply to any of the higher-order invariants constructable from the Riemann tensor. The classical method of Lagrange multipliers thus proves a powerful tool in analyzing the equations of motion for higher-order Lagrangians.

This preliminary investigation of the quadratic Lagrangians poses several interesting problems and lines of inquiry. What is the physical significance and interpretation of the nonvanishing Lagrange multiplier constraint forces and the "additional" terms to which they give rise in the equations of motion in the 4-formalism of Sec. II? What is the significance of the fact that these constraints are automatically satisfied for the Lagrangian $g^{1/2}R$ of general relativity? Is general relativity unique in this regard? (Recall that in Stephenson's cases A and B, Higgs has already shown that the assumption $\Lambda_\alpha{}^{\beta\gamma} = 0$ reduces these theories to general relativity.)

One would also like to have the remaining quadratic invariants $R_{\mu\nu}R^{\mu\nu}$ and $R_{\alpha\beta\gamma\delta}R^{\alpha\beta\gamma\delta}$ in 3 + 1 Hamiltonian form. In principle the application of the "Lagrange multiplier" method developed in Sec. III to these invariants should be straightforward and direct; however, it is obvious that in the (3 + 1)-formalism the calculations involved for higher-order Lagrangians will rapidly become very tedious and elaborate. It is hoped that the present paper will stimulate interest in the ADM formalism in the context of higher-order Riemann invariants.

[1]C.W. Misner, K.S. Thorne, and J.A. Wheeler, *Gravitation* (Freeman, San Francisco, 1973), Chap. 21.

[2]G. Stephenson, Nuovo Cimento 9, 263 (1958).

[3]P. Higgs, Nuovo Cimento 11, 816 (1959).

[4]I.W.Roxburgh, J. Phys. A 8, 1075—80 (1975).

[5]C. Lanczos, Rev. Mod. Phys. 29, 337 (1957).

[6]J.R. Ray, Nuovo Cimento B 25, 706 (1975).

[7]F.C. Michel, Ann. Phys. (N.Y.) 76, 281 (1973).

[8]E. Pechlaner and R. Sexl, Commun. Math. Phys. 2, 165 (1966).

[9]B. DeWitt, in *Relativity. Groups and Topology*, edited by C. DeWitt and B. DeWitt (Gordon and Breach, New York, 1962).

[10]C. Lanczos, Rev. Mod. Phys. 34, 379 (1962).

[11]R. Adler, M. Bazin, and M. Schiffer, *Introduction to General Relativity* (McGraw-Hill, New York, 1965), Chap. 2.

[12]H. Buchdahl, Proc. Camb. Phil. Soc. 56, 398 (1960).

[13]R. Arnowitt, S. Deser, and C.W. Misner in *Gravitation: An Introduction to Current Research* (Wiley, New York, 1962), Chap. 7.

1537   J. Math. Phys., Vol. 17, No. 8, August 1976

J.L. Safko and F. Elston   1537

Quantum mechanics on topological networks

Aaron B. Budgor

Department of Chemistry, University of California, La Jolla, California 92037
(Received 11 September 1975)

A quantum mechanical model of electrons on a network which has been used to obtain the line spectra of conjugated organic molecules and the band structure for various solid state systems is shown to be intimately connected with the homology and cohomology sequences of the algebraic topology of 1-complexes. Thus, the remarkable qualitative agreement of these models to experiment implies that those calculated physical quantities are structural quantities. Quantum mechanical conservation laws analogous to Kirchhoff's electrical conservation laws are derived by variational methods applied to the system's Lagrangian. The topology is shown to enforce energy conservation on the network.

1. INTRODUCTION

Network modeling of physical systems has found widespread acceptance in science. To chemists and physicists this manifests itself mostly in a variety of discretized models of molecules and solids in which a three-dimensional material is idealized as atoms connected by bonds lying in R^3. Modern day mechanical and electrical engineers have also found network modeling very beneficial in dealing with aggregates of complex systems. They, however, utilize a different class of networks which we shall for now term "Kirchhoff-like," after G. Kirchhoff. These networks have the feature of inherently incorporating conservation principles, energy for example, as part of their algebraic structure. Although electrical networks are the best example of this class, there is a vast literature on electrical circuit analogies to the electromagnetic, elastic, and fluid flow field equations and to mechanical and hydraulic pipe flow systems.[1,2] A number of publications have also appeared in the quantum mechanical literature describing a Kirchhoff-like network model of electrons in molecules and solids.[3-8] The remarkable success all these models have had in emulating the dynamic and kinematic features of the systems they represent seems to indicate some deep underlying mathematical structure common to all. Due to a very general theorem in electrical circuit theory, Tellegen's theorem, the existence of such structure is even more compelling. It has been used to derive virtually every theorem concerning the power distribution in an electrical circuit, linear or nonlinear, of arbitrary topology.[9]

We define an electrical circuit as a graph consisting of vertices connected to each other by directed branches on which Kirchhoff's current and voltage laws are obeyed. The current law (KCL) states that at any instant of time the net current into each vertex is zero, while the voltage ˙ ˙ (KVL) states that instantaneously the voltage arou ˌch loop of the circuit is zero regardless of wheu.ₓᵣ the constitutive processes occuring on the circuit are linear or nonlinear, passive or active, singly or multiply valued. The excitation is also arbitrary as it may be sinusoidal, exponential, periodic, or random. It is well-established that both (KCL) and (KVL) are topological in origin and that Tellegen's theorem is of natural consequence in the algebraic topology of 1-complexes.[10-12] In brief, (KCL) is the

electrical counterpart of what the topologist calls a homology sequence, while (KVL) is the cohomology sequence. An Ohms law relation comes about as an isomorphism between the two sequences and Tellegen's theorem as a consequence of their orthogonality (in the sense of a vector space). For higher-dimensional topological complexes it can be shown that the operational structures of linear graph theory and the vector calculus are identical. This is the dominant reason for the great success of network analogies to problems in continuum field theories. More importantly, however, is their possible utility as a foundation for the theory of nonlinear equations of motion, of which the Brayton—Moser equations are one example.[13]

In this paper we restrict ourselves to a discussion of the aforementioned one-dimensional (1-complex) network model of electrons in molecules and solids. Generalizations to higher-dimensional complexes and their relationships to continuum quantum fields will be considered elsewhere. Originated as a free electron model to describe in a simple fashion the electronic properties of conjugated organic molecules[3] and molecular solids such as graphite and diamond,[4,5] it was later extended to nonzero atomic potentials by Montroll and his colleagues and utilized to analytically determine the band structure of perfect crystals, crystals with defects, and crystal surfaces.[6,7] Later applications were directed towards electron scattering by crystal defects.[8] In this model electrons are restricted by nonconstant one-dimensional potential fields to move along the one-dimensional bonds of a network of atoms. The network is given the same topological pattern with the same lattice spacings that would represent its molecular or crystallographic structure. Energy spectra are then obtained by incorporating the boundary conditions that

(i) the wavefunction is continuous along all branches (bonds) of the network and at the node points, and

(ii) the net current flowing away from any node point must vanish.

The technical details of this procedure are discussed at great length in Refs. 3–8. It should be noted that the first condition only holds true for atoms with dimensions of measure zero, i.e., atoms represented as points. (ii) has the flavor of a "Kirchhoff-like" current law. It will be shown in Sec. 3 that in the stationary

state the conservation condition (ii) has, for an arbitrary potential $V(x)$, $x \in R$, the form

$$\sum_{\{j_p\}} \frac{d\psi(x_{jj_p})}{dx_{jj_p}} \bigg|_{x_{jj_p} = \text{node}} = 0, \quad \psi \in L_2[j, j_p], \quad (1.1)$$

where $\psi(x_{jj_p})$ is the Schrödinger wavefunction with support on $[j, j_p]$ and the sum in (1.1) is over all connected lattice points j_p to j.

Comparisons of the energy spectra obtained with this method to experiment and to the more accepted theoretical analyses has shown a remarkable accuracy considering the one-electron approximation used and the one-dimensionality of the model.[3,4]

In the forthcoming section we shall briefly review the relationship of Tellegen's theorem to the algebraic topology of 1-complexes. The consequences of these generalized topological (KCL) and (KVL) variables will then be exploited in Secs. 3 and 4. In Sec. 3 this is begun with a general discussion on the embedding of a kinematic constitutive conservation theorem, the conservation of momentum, on a topological structure. These results are specialized to derive the quantum mechanical (KCL) condition, Eq. (1.1), by a variational procedure. It will be shown (Sec. 4) that this condition implicitly contains the information of (KVL) and results in a net energy conservation for the aggregate network. By virtue of this superposition of physics on mathematical structure, the calculable physical quantities are therefore structural quantities.

2. TELLEGEN'S THEOREM AND THE ALGEBRAIC TOPOLOGY OF NETWORKS

Tellegen's theorem may be written as

$$\sum_p \Lambda' i_p \Lambda'' v_p = \sum_b \Lambda' i_b \Lambda'' v_b, \quad (2.1)$$

where $i_{p(b)}$ and $v_{p(b)}$ are, respectively, the current and voltage as measured along any branch of the circuit, with p denoting a port (open branch) and with b denoting an internal branch. The sums are performed over all the ports and internal branches in the circuit (see Fig. 1). Λ' and Λ'' are referred to as Kirchhoff's operators and have the following properties:

(1) If Kirchhoff's laws for currents and/or voltages are valid for a circuit of particular topology, then they will also have validity after operation by a Kirchhoff operator.

(2) All linear operators and *both* Kirchhoff current and voltage operators.

(a)   (b)
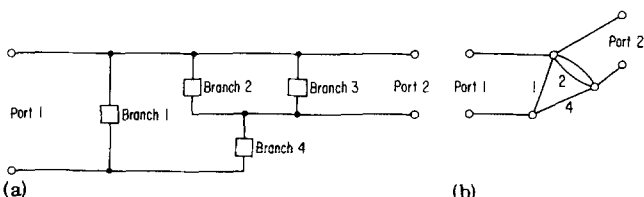
FIG. 1. Typical network with four internal branches and two ports: (a) circuit diagram showing the elements, as yet unspecified, and the branches and ports numbered; (b) topology of this network showing the branches and ports numbered.

TABLE I. Kirchhoff operators.

Kirchhoff current (voltage) operators are defined as those which yield, from a set of currents (voltages) that obey Kirchhoff's current (voltage) law, a set of numbers or functions that also obey Kirchhoff's current (voltage) law. The resulting quantities need not have the dimensions of current (voltage) and may depend upon other parameters or variables (such as frequency or temperature) introduced by the operator. All linear operators (that operate in the same way on all branches and ports of the network) are Kirchhoff operators. Some examples of linear operators are the following (these are all Kirchhoff operators):

1. Identity: $\Lambda i = i(t)$
2. Multiplication by a constant or by a specific function of time $f(t)$: $\Lambda i = f(t) i(t)$
3. Shift in time by t_0: $\Lambda i = i(t - t_0)$
4. Differentiation in time: $\Lambda i = di(t)/dt$
5. Integration in time: $\Lambda i = \int i(\tau) d\tau$
6. Convolution with a specific function of time $f(t)$: $\Lambda i = \int_{-\infty}^{\infty} i(t - \tau) f(\tau) d\tau$
7. Evaluation of i for a specific time t_0: $\Lambda i = i(t_0)$
8. Time reversal: $\Lambda i = i(-t)$
9. Selection of the even (or odd) part of $i(t)$: $\Lambda i = \frac{1}{2}[i(t) + i(-t)]$
10. Time average (or stochastic average of an ergodic process): $\Lambda i = \overline{i(t)}$
11. Selection of first-order perturbations or, more generally, nth order perturbations
12. Selection of a particular experiment; the various experiments may involve different element values of different excitations, but they always involve the same topology.
13. Taking the Fourier or Laplace transorm (or, for periodic signals, selection of the Fourier coefficients)
14. Complex conjugation

(3) The nonlinear operators that presently qualify for the title have been found empirically to be either current operators or voltage operators, but never both simultaneously.

A compilation of proper and improper Kirchhoff operators has been listed in Tables I and II. In particular, if $\Lambda'' = \Lambda' = I$, then (2.1) is nothing more than a power conservation theorem which states that the power flowing in and out of an open system is the same as the internal power generated in the system. If the system has no ports, i.e., it is closed, then the lhs of (2.1) vanishes, and

$$\sum_b \Lambda' i_b \Lambda'' v_b = 0. \quad (2.2)$$

As a more interesting example of (2.1) consider the two circuits depicted in Figs. 2 and 3. Although they

TABLE II. Non-Kirchhoff operators.

Some examples of operators that are *not* Kirchhoff operators are the following:

1. Squaring: $\Lambda i = i^2(t)$
2. Taking the absolute value: $\Lambda i = |i(t)|$
3. Determination of the maximum value in a certain time range
4. Selection of the root-mean-square or effective values
5. Selection of amplitude-modulation or frequency-modulation components
6. Multiplication by constants or functions of time that are different for different branches
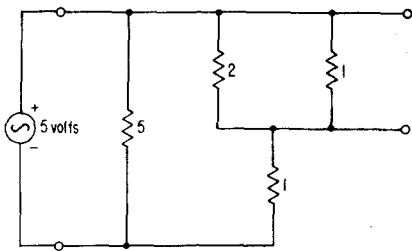
FIG. 2. One state of the network (values of resistances in ohms).

both have the same topology different constitutive relations occur on them. By associating state 1 with Λ' and state 2 with Λ'' we obtain item no. 12 in Table I. Equation (2.1) is a reciprocity relation because the two states of the network are related solely by their topology and can be likened to the occurrence of a virtual process since state 2 does not occur on circuit 1 and vice-versa.

We shall now demonstrate the relation between Tellegen's theorem and algebraic topology. The theorems and corollaries to be stated are given without proof, although the references where they may be found are provided.[11,12,14] Definitions for the topological terms used are given in Appendix A. For the sake of clarity it should be noted that, although Tellegen's theorem has been applied to electrical networks embedded in R^2 one would like to show its applicability to molecules and to solids in R^3. This can be accomplished by utilizing the following graph projection theorem. For notational simplicity graph is synonymous with linear graph.

Theorem 2.1:$\forall$ $\Gamma \subset R^3 \exists$ a graph isomorphism Π: $\Gamma \to \Sigma$ onto a two-dimensional orientable manifold $\Sigma \subset R^3$ of genus g.

The map Π must be performed such that no two edges in Σ cut one another except at the vertices.

As an illustration of this theorem, Euler has shown that if Γ is a convex polyhedron with n vertices, m edges, and f finite faces, then Σ is a planar graph and $n - m + f = 1$. The f faces form a basis of elementary cycles. When the mapping Π does not yield a planar graph, the number of elements in the basis of elementary cycles, $k(\Sigma)$, is[14]

$$k(\Sigma) = (f - 1) + r(\Sigma), \quad \text{where } 0 \leqslant r(\Sigma) \leqslant 2g. \quad (2.3)$$

Thus, the image of Γ partitions the surface of Σ into $m + 2 - n - r(\Sigma)$ faces, with each face homeomorphic to $I^2 = [0,1] \times [0,1]$. These examples suggest that no matter what topology Γ has, one can always find a two-dimensional manifold in R^3 in which Σ is a 1-chain.

The following definitions and theorems will now be introduced in order to motivate the central result of this section, Theorem 2.4.

Definition: Let Σ be a graph consisting of m arcs $\hat{\sigma}_1^1, \ldots, \hat{\sigma}_m^1$, then a flow is a vector $\phi = (\phi_1, \ldots, \phi_m)$ such that

(a) for every $k \leqslant m$, $\phi_k \in R$, and ϕ_k is denoted as the flow in arc $\hat{\sigma}_k^1$, (2.4)

(b) for every vertex a $\sum\limits_{i \in \omega^-(a)} \phi_i = \sum\limits_{i \in \omega^+(a)} \phi_i$.

Theorem 2.2: A necessary and sufficient condition for a vector ϕ to be a flow is if it is of the form

$$\phi = \sum_{i=1}^{m} s_i \mu^i, \quad (2.5)$$

where the $s_i \in R$, $i = 1, \ldots, m$, and the μ^k are elementary cycles, $k = 1, \ldots, m$. Thus, ϕ is a 1-cycle and is orthogonal to all 1-coboundaries.

Definition: A potential difference (tension) is a vector $\theta = (\theta_1, \ldots, \theta_m)$ such that for every elementary cycle

$$\sum_{i \in \mu^+} \theta - \sum_{i \in \mu^-} \theta = 0. \quad (2.6)$$

Theorem 2.3: A vector θ is a potential difference if $\exists$ a function $t(a)$ defined on a set X of vertices with values in R such that for every arc $i = (a, b)$, $\theta_i = t(b) - t(a)$. Thus, $t(a)$ is denoted as the potential associated with θ.

Corollary 2.3: A necessary and sufficient condition for a vector θ to be a potential difference is that it be of the form

$$\theta = \sum_{i=1}^{m} s_i \omega^i, \quad (2.7)$$

where the ω^i are elementary coboundaries and $s_1, \ldots, s_m \in R$. θ, therefore, is a 1-coboundary and is orthogonal to any 1-cycle. Finally,

Theorem 2.4: A vector $\phi \in R^m$ is a flow iff it is orthogonal to every vector in Θ, the set of all potential differences; a vector $\theta \in R^m$ is a potential difference iff it is orthogonal to every vector in Φ, the set of all flows.

One may conclude from this theorem that Φ and Θ are orthogonal subspaces of R^m, that is, $\Phi \oplus \Theta = R^m$. Thus, the inner product of every ϕ and θ vanishes:

$$\langle \phi, \theta \rangle = 0. \quad (2.8)$$

If we now let Λ' and Λ'' be linear operators that leave the vector space of flows and tensions invariant, then by premultiplying (2.5) and (2.7) by Λ' and Λ'', respectively, and then taking the inner product as in (2.8) we have

$$\langle \Lambda'\phi, \Lambda''\theta \rangle = 0. \quad (2.9)$$

The proof of (2.9) is easily demonstrated. Since ϕ and θ are 1-cycles and 1-coboundaries, respectively, then by linearity so are $\phi' = \Lambda'\phi$ and $\theta' = \Lambda''\theta$. Application of Theorem A3.1 completes the proof.
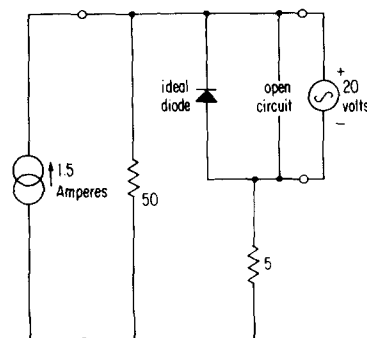
FIG. 3. Another state of the network (values of resistances in ohms).

Our correspondence of algebraic topology with
Tellegen's theorem is now complete by noting the sim-
ilarity of (2.4) with the Kirchhoff current law, of (2.6)
with the Kirchhoff voltage law, and of (2.9) with
Tellegen's theorem, Eq. (2.1). These results encom-
pass the case of port branches and open circuits since
"open branch" flows and tensions of the form (2.5) and
(2.7) may be constructed by including ideal arcs of null
contribution to the elementary cycles and coboundaries
of such ϕ and θ.

3. FLOW CONSERVATION ON A NETWORK

The flow and potential difference properties of a
linear graph when identified with a physical process
are, in principle, conservative. This is clearly ap-
parent from the Kirchhoff and d'Alembert "principle of
virtual work"[15] laws, the by-products of graphical
analogies to electrical and mechanical systems. In
general, conservation laws can be elegantly obtained
through the methods of variational calculus when applied
to a Lagrangian. This approach still holds true for
systems represented by graphs, with the sole exception
that the domain of the Lagrangian is now restricted to
a 1-chain. We now introduce two variational procedures
which will subsequently be used to derive flow 1-chains
in quantum field theory.

Consider the well-known classical mechanics contin-
uity equation

$$\frac{\partial \rho}{\partial t} + \nabla \cdot (\rho v) = 0 \tag{3.1}$$

and its quantum analog

$$\frac{\partial}{\partial t}(\psi^* \psi) + \frac{i\hbar \nabla}{2m} \cdot (\psi \nabla \psi^* - \psi^* \nabla \psi) = 0. \tag{3.2}$$

We define the scalar functions ρ and $\psi\psi^*$ in R^{3N} as the
particle (fluid) density and probability density, respec-
tively, and the vectors ρv and $i\hbar(\psi \nabla \psi^* - \psi^* \nabla \psi)/2m$ as
currents.

Let M be a differentiable manifold with tangent bundle
$T(M)$

$$M = R^3 \times R^3 \times \overset{6N-9}{\cdots\cdots\cdots} \times R^3 \times R^2 \tag{3.3}$$

and consider the Lagrangian function $L(\phi, \phi_q)$ on $T(M)$
where ϕ is a $(3N+1)$-dimensional field with coordinates
$q_1, \ldots, q_{3N+1}$. Equation (3.1) results from a variation
of the action integral

$$A = \int L dq^{3N+1}. \tag{3.4}$$

Denoting $p_\alpha = \partial L / \partial \phi_{q_\alpha}$, $\alpha = 1, \ldots, 3N+1$, the Euler–
Lagrange equation associated with (3.4) is

$$\frac{\partial L}{\partial \phi} - \sum_{\alpha=1}^{3N+1} \frac{\partial p_\alpha}{\partial q_\alpha} = 0. \tag{3.5}$$

Suppose ϕ is kinosthenic, that is, a variable not ex-
plicitly contained in L, then

$$\frac{\partial L}{\partial \phi} = 0 \tag{3.6a}$$

and (3.5) becomes

$$\sum_{\alpha=1}^{3N+1} \frac{\partial p_\alpha}{\partial p_\alpha} = 0. \tag{3.6b}$$

With the correspondence $(q_{i\bmod(3)} = x_i,\ q_{(i+1)\bmod(3)}$
$= y_i,\ q_{(i+2)\bmod(3)} = z_i, q_{3N+1} = t)$ and $(p_{i\bmod(3)} = v_{x_i},$
$p_{(i+1)\bmod(3)} = v_{y_i},\ p_{(i+2)\bmod(3)} = v_{z_i},\ p_{3N+1} = m),\ i, \ldots, N,$
we arrive at (3.1).

Equation (3.2) can be obtained by an application of
Noether's principle to the N-particle quantum mechani-
cal Lagrangian

$$L = \psi^*(\mathbf{x}_1, \ldots, \mathbf{x}_N) \left(\frac{\hbar}{i} \frac{\partial \psi}{\partial t} (\mathbf{x}_1, \ldots, \mathbf{x}_N) \right.$$
$$+ V(\mathbf{x}_1, \ldots, \mathbf{x}_N) \psi(\mathbf{x}_1, \ldots, \mathbf{x}_N) \bigg) \tag{3.7}$$
$$+ \frac{\hbar^2}{2m} \nabla \psi^*(\mathbf{x}_1, \ldots, \mathbf{x}_N) \cdot \nabla \psi(\mathbf{x}_1, \ldots, \mathbf{x}_N), \quad \mathbf{x}_i \in R^3,$$

where $(\mathbf{x}_1, \ldots, \mathbf{x}_N)$ are the N-particle coordinates,

$$\nabla = \sum_{\alpha=1}^{N} \frac{\partial}{\partial \mathbf{x}_\alpha},$$

and $V(\mathbf{x}_1, \ldots, \mathbf{x}_N)$ is the N-particle interaction potential
and/or an external electric field. This variational ap-
proach makes use of the property that the action integral
remains invariant with respect to a group of infinitesi-
mal transformations applied either to the dependent or
independent variables. Since this Lagrangian is invari-
ant up to an arbitrary complex phase factor, $i\alpha(\mathbf{x}, t)$,
$\alpha(\mathbf{x}, t)$ being infinitesimal, then by substituting
$\exp[i\alpha(\mathbf{x}, t)]\psi$ for ψ (3.7) becomes

$$\bar{L}(\mathbf{x}_1, \ldots, \mathbf{x}_N, \alpha)$$
$$= L + \hbar \psi^*(\mathbf{x}_1, \ldots, \mathbf{x}_N) \psi(\mathbf{x}_1, \ldots, \mathbf{x}_N) \frac{\partial \alpha}{\partial t}$$
$$+ \frac{\hbar^2 i}{2m} \left(\sum_j \psi(\mathbf{x}_1, \ldots, \mathbf{x}_N) \frac{\partial \psi^*}{\partial \mathbf{x}_j}(\mathbf{x}_1, \ldots, \mathbf{x}_N) \frac{\partial \alpha}{\partial \mathbf{x}_j} \right. \tag{3.8}$$
$$\left. - \psi^*(\mathbf{x}_1, \ldots, \mathbf{x}_N) \frac{\partial \psi}{\partial \mathbf{x}_j}(\mathbf{x}_1, \ldots, \mathbf{x}_N) \frac{\partial \alpha}{\partial \mathbf{x}_j} \right) + O(\alpha^2).$$

Variation of (3.8) with respect the kinosthenic vari-
able α yields the probability density conservation law.

Relations (3.1) and (3.2) are also valid when evaluated
at the vertices of a graph. We verify this for (3.1); by
analogy it holds true for (3.2). For simplicity we as-
sume temporal invariance, that is, the time dimension
does not enter in as a directed arc in the graph.

As a first step we rewrite (3.1) as a conservation of
flux over the submanifold $D \subset M$

$$\frac{\partial}{\partial t} \int_D \rho \, dx^{3N} + \int_D \nabla \cdot (\rho v) \, dx^{3N} = 0. \tag{3.9}$$
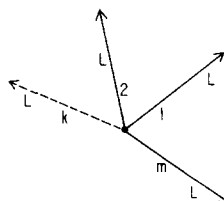
FIG. 4. A node with m di-
rected bonds of length L em-
anating from it.

By Gauss' theorem the second integral can then be related to a $(3N-1)$-dimensional integral over the boundary of D:

$$\int_D \nabla \cdot (\rho v)\, dx^{3N} = \int_{\partial D} \rho v\, dx^{3N-1} = 0. \tag{3.10}$$

Suppose D is a 1-chain $(D \in R^N)$, and without loss of generality consider one central node with m attached arcs, each of length L (Fig. 4). Then

$$D = \sum_{j=1}^{m} a_j^1 \hat{\sigma}_j^1 \tag{3.11a}$$

and

$$\partial_1 D = \sum_{j=1}^{m} a_j^1 \partial_1 \hat{\sigma}_j^1. \tag{3.11b}$$

Since an arbitrary directed arc $\hat{\sigma}_i^1$, connecting the central node n_0 and some node n_i, can be represented as

$$\hat{\sigma}_i^1 = (n_0, n_i) \tag{3.12a}$$

then

$$\partial_1 \hat{\sigma}_i^1 = \hat{\sigma}_i^0 = n_i - n_0. \tag{3.12b}$$

Thus, the current 1-chain ρv is conserved at 0-cells $\hat{\sigma}_j^0$.

For many physical situations it is unimportant to know the time development of ψ. It is therefore of interest to determine a stationary state conservation condition for real ψ. Defining $\psi_{(x_1,\ldots,x_N)}^{(j)}$ as the N-particle wavefunction on bond j, $1 \le j \le m$, variation of the action integral A gives

$$\delta A = \delta \int_0^L L(\psi_{(x)}^{(1)},\ldots,\psi_{(x)}^{(m)},\ \psi_{(x)}'^{(1)},\ldots,\psi_{(x)}'^{(m)},x)\,dx^{(j)}$$

$$= \sum_{j=1}^{m} \int_0^L \left(\frac{\partial L}{\partial \psi_{(x)}^{(j)}}\,\delta\psi_{(x)}^{(j)} + \frac{\partial L}{\partial \psi_{(x)}'^{(j)}}\,\delta\psi_{(x)}'^{(j)} \right) dx^{(j)} \tag{3.13}$$

$$= \sum_{j=1}^{m} \left[\frac{\partial L}{\partial \psi_{(x)}'^{(j)}}\,\delta\psi_{(x)}^{(j)} \Big|_0^L \right.$$

$$\left. + \int_0^L \left(\frac{\partial L}{\partial \psi_{(x)}^{(j)}} - \frac{d}{dx^{(j)}} \frac{\partial L}{\partial \psi_{(x)}'^{(j)}} \right) \delta\psi_{(x)}^{(j)}\,dx^{(j)} \right],$$

where L is (Appendix B),

$$L = \frac{\hbar^2}{2m} \sum_{i=1}^{N} \left(\frac{\partial \psi(x_1,\ldots,x_N)}{\partial x_i} \right)^2 \tag{3.14}$$

$$+ [V(x_1,\ldots,x_N) - E]\psi^2(x_1,\ldots,x_N)$$

and where we have employed the notational substitution

$$(x) = (x_1,\ldots,x_N).$$

Substitution of (3.14) into (3.13) yields at the node point $x_i = 0$

$$\sum_{j=1}^{m} \sum_{i=1}^{N} \left(-\frac{\hbar^2}{m} \frac{\partial\psi_{(x)}^{(j)}}{\partial x_i^{(j)}} \Big|_{x_i^{(j)}=0} \delta\psi_{(0)}^{(j)} \right.$$

$$\left. + \int_0^L \left\{ 2[V(x) - E]\psi_{(x)}^{(j)} - \frac{\hbar^2}{m}\,\Delta\psi_{(x)}^{(j)} \right\} \delta\psi_{(x)}^{(j)}\,dx^{(j)} \right) = 0. \tag{3.15}$$

In order for the equality in (3.15) to hold, each term must vanish separately. Since $\delta\psi_{(0)}^{(j)}$ and $\delta\psi_{(x)}^{(j)}$ in the first and second terms are arbitrary functions in the domain

of definition of x, then the conservation condition is

$$\sum_{j=1}^{m} \sum_{i=1}^{N} \frac{\partial\psi_{(x)}^{(j)}}{\partial x_i^{(j)}} = 0. \tag{3.16}$$

To exhibit the complete generality of the conservation conditions (3.2) and (3.16), consider the one-particle Klein–Gordon equation in the presence of an electromagnetic field,

$$\left\{ \sum_\mu [\hat{P}_\mu - (e/c)A_\mu]^2 + m^2c^2 \right\} \psi = 0, \tag{3.17}$$

where P_μ is the 4-vector $P_\mu = (P_1, P_2, P_3,\ ieA_0/c)$ with components $P_\beta = (\hbar/i)\partial/\partial x_\beta$ and where $x_\mu = (x,y,z,ict)$. This equation can be derived from the Lagrangian

$$L = -\frac{1}{c^2} \frac{\partial\psi^*}{\partial t} \frac{\partial\psi}{\partial t} + \nabla\psi^* \cdot \nabla\psi - \frac{2e\psi^*}{\hbar ci} \left(A \cdot \nabla\psi + \frac{A_0}{c}\frac{\partial\psi}{\partial t} \right) \tag{3.18a}$$

$$+ \frac{e^2}{\hbar^2 c^2}(A^2 - A_0^2 + m^2c^4)\psi^*\psi$$

and implicitly contains the gauge invariance of the vector potential

$$\frac{1}{c}\frac{\partial A_0}{\partial t} + \nabla \cdot A = 0. \tag{3.19}$$

Making the transformation $\bar\psi = \exp[i\alpha(x,t)]$, (3.18a) becomes

$$\bar{L} = L - \frac{i}{c^2}\frac{\partial\alpha}{\partial t} \left(\psi\frac{\partial\psi^*}{\partial t} - \psi^*\frac{\partial\psi}{\partial t} + \frac{2e}{\hbar i}A_0\psi\psi^* \right)$$

$$+ i\nabla\alpha \cdot \left(\psi\nabla\psi^* - \psi^*\nabla\psi - \frac{2e}{\hbar ci}A\psi\psi^* \right) + O(\alpha^2). \tag{3.18b}$$

Variation with respect to the kinosthenic variable $\alpha(x,t)$ results in the current conservation condition

$$\sum_\mu \frac{\partial j_\mu}{\partial x_\mu} = 0 \tag{3.20a}$$

with $j_\mu = (j_1, j_2, j_3, ic\rho)$ and where

$$\rho = \frac{ie\hbar}{2mc^2} \left(\psi^*\frac{\partial\psi}{\partial t} - \psi\frac{\partial\psi^*}{\partial t} \right) - \frac{e^2 A_0\psi\psi^*}{mc^2}, \tag{3.20b}$$

$$j_\beta = \frac{e\hbar}{2mi} \left(\psi^*\frac{\partial\psi}{\partial x_\beta} - \psi\frac{\partial\psi^*}{\partial x_\beta} \right) - \frac{e^2 A\psi\psi^*}{mc}, \quad \beta = 1,2,3. \tag{3.20c}$$

Equations (3.20) also apply for an N-particle system, each particle having its own local coordinate x_i.

On a graph the conservation condition (3.20a) is now a flow 1-chain, $j_\mu \in R$, and must be evaluated at vertices. If the 1-chain is a combination of time and space 1-cells the appropriate current component (3.20b) or (3.20c) must be used. Thus, vertices to which both time and space 1-cells are connected will have a mixed conservation condition.

In the absence of an electromagnetic field in (3.17), $A_\mu = 0$, the conservation condition (3.16) is then applicable when ψ is a time independent real scalar field.

In conclusion, it should be mentioned that all the conservation conditions previously derived apply equally well upon replacement of the scalar field ψ by quantized field operators $\tilde\psi$. In such a case the current conserva-

tion condition on a graph becomes

$$\sum_i j(x)\big|_{x=\text{vertex}} = \sum \lim_{\substack{t' \to t \\ x' \to x^+}} \left(\frac{\partial}{\partial x} - \frac{\partial}{\partial x'}\right) G(x,t;x',t') = 0$$

(3.21)

with the sum performed over all arcs connected to vertex x. $G(x,t;x',t')$ is the time or temperature dependent one particle Green's function.[16]

4. ENERGY CONSERVATION ON A NETWORK

In Sec. 2 Tellegen's theorem was introduced as a statement concerning the conservative nature of energy (power) on a network. For an electrical circuit this energy comes from the product of a (current)×(voltage), each factor being intimately connected with the topological structure of the network. To carry this analogy over to network models of other physical systems, one must obtain an identification of two canonical variables whose properties are those of a flow and a potential difference.

For mechanical systems this identification is simple, current $\leftrightarrow$ velocity, and because of Newton's third Law, voltage $\leftrightarrow$ force. The product of these two variables has dimensions of energy. In quantum mechanics this statement can be made explicit by rewriting the current in (3.2) as

$$\mathbf{j} = (1/m)\,\mathrm{Re}(\nabla\psi/i) = \mathbf{P}/m = \mathbf{v}.$$

(4.1)

(4.1) is clearly a (KCL) variable.

We choose the wavefunction ψ as our quantum mechanical "force." This has the appropriate form since the solution of the one-dimensional stationary Schrödinger equation can be written as

$$\psi(x) = \psi_h(x) + (1/k)\int_0^x \sin k(x-x')V'(x')\psi(x')\,dx', \quad (4.2a)$$

where

$$k^2 = 2mE/\hbar^2, \quad V'(x) = (2m/\hbar^2)V(x)$$

(4.3)

and $\psi_h(x)$ is the homogeneous solution in the absence of the potential field $V(x)$.

Integration of the integral in (4.2a) by parts yields

$$\psi(x) = \psi_h(x) + \frac{1}{k^2}\left[V'(x)\psi(x) - \cos kx V'(0)\psi(0)\right]$$

$$- \frac{1}{k^2}\int_0^x \cos k(x-x')\frac{d}{dx'}(V'(x')\psi(x'))\,dx'$$

$$= \psi_h(x) + \frac{1}{k^2}\left[V'(x)\psi(x) - \cos kx V'(0)\psi(0)\right]$$

$$+ \frac{1}{k^2}\int_0^x \cos k(x-x')F_{\text{Q.M.}}(x')\,dx'.$$

(4.2b)

Since $\psi(x)$ is continuous at vertex points, it is truly a (KVL) variable.

Energy conservation is then derived by application of Tellegen's theorem, which for Fig. 4 results in

$$\sum_{\alpha=1}^m \psi_\alpha(0)\frac{d\psi(x_\alpha=0)}{dx_\alpha} = 0$$

(4.4)

Continuity of all the ψ_α at $\psi_\alpha = 0$ reduces (4.4) to the

form (1.1), which is then the starting point from which the line and band spectra of molecules and solids are derived.

APPENDIX A

In this appendix we define, for the reader's convenience, a number of topological concepts necessary in the development of the main text. Due to the extensive literature on this subject, proofs will be omitted.

Definition: A linear graph Γ is an interconnected set of arcs (1-cells, $\hat{\sigma}_j^1$) at vertices (0-cells, $\hat{\sigma}_j^0$).

If the arcs are directed then Γ is said to be *orientable* if it is possible to define a positive rotation at each of its points, all the rotations being consistent. For nonlinear graphs we generalize the concept of orientation by introducing the genus of a surface.

Definition: An orientable surface is said to be of *genus g* if by elastic deformations it can be made to coincide with a sphere of g handles: The ordinary sphere is of genus 0, the torus of genus 1, and the pretzel of genus 2.

Γ can be given vector space structure by defining linear combinations, k-chains, of the k-cells, $k = 0,1$.

Definition: A k-chain, $\hat{c}^k$, on a graph Γ with n vertices and m arcs is some linear combination using the k-cells $\hat{\sigma}_j^k$ as a basis:

$$\hat{c}^k = \sum_j a_j^k \hat{\sigma}_j^k, \quad a_j^k \in R.$$

(A1)

Since the set of all k-cells is a vector space V^k, $\hat{c}^k \in V^k$ and

$$\dim V^1 = m, \quad \dim V^0 = n.$$

(A2)

Definition: The *boundary operator* ∂_k is a linear operator

$$\partial_k: V^k \to V^{k-1}$$

(A3a)

with the fundamental property

$$\partial_{k-1} \circ \partial_k(\cdot) = 0.$$

(A4)

In words, (A4) states that the boundary of the boundary of a k-chain vanishes. As an example of (A3a) the boundary of the arc (a,b) is

$$\partial_1(a,b) = b - a.$$

(A3b)

Chains whose boundaries are zero are called *k-cycles*. The set of all k-cycles, $\gamma^k = \ker(\partial_k)$, is a vector subspace of V^k, $Y^k \subset V^k$.

As an illustration a 1-cycle is a sequence of arcs $(\hat{\sigma}_1^1,\ldots,\hat{\sigma}_q^1)$, $\hat{\sigma}_j^1 \in V^1$, such that

(a) every arc $\hat{\sigma}_k^1$, $k = 1,\ldots,q$, is joined to the preceding arc $\hat{\sigma}_{k-1}^1$ at one of its extremities and to the succeeding arc $\hat{\sigma}_{k+1}^1$ at the other (it is a 1-chain),

(b) the sequence does not use the same arc twice,

(c) the initial and terminal vertices of the 1-chain coincide.

An *elementary cycle* satisfies in addition to the previous three conditions,

(d) in traversing the 1-cycle one encounters the same vertex (excluding the starting point) only once.

For a given 1-cycle μ, we denote by μ^+ the set of arcs oriented in a given sense, and by μ^- the set of arcs oriented in the opposite sense. A vector $\mu = (\mu_1, \ldots, \mu_m)$ can be associated with every 1-cycle of a graph consisting of m arcs in such a way that

$$\mu_i = \begin{pmatrix} 0 & \text{if } i \notin \mu^+ U \mu^- \\ 1 & \text{if } i \in \mu^+ \\ -1 & \text{if } i \in \mu^- \end{pmatrix}. \tag{A5}$$

Theorem A.1: Every 1-cycle is a direct sum of elementary cycles. Since the V^k are linear vector spaces, we can define a scalar product

$$\langle \cdot, \cdot \rangle : V^{k*} \times V^k \to R \tag{A6a}$$

$$\langle \hat{c}^{k*}, \hat{c}^k \rangle = \sum_j a_j^{k*} a_j^k, \tag{A6b}$$

where $\hat{c}^{k*}$ is a k-*cochain* defined by the linear mapping $V^{2-k} \to R$ and is dual to $\hat{c}^k$.

The collection of all k-cochains forms an m-dimensional vector space V^{k*}. Geometrically the dual graph Γ^* has $n^* = f$ vertices and $m^* = m$ arcs which partition Γ^* into $f^* = n$ faces [*Remark*: Each vertex of Γ lies in a unique and separate face of Γ^*.]

The adjoint to ∂_k, δ_{k-1}, is a *coboundary operator* defining a linear map

$$\delta_{k-1} : V^{k*-1} \to V^{k*} \tag{A7}$$

and obeying the property

$$\langle \partial_k \hat{c}^k, \hat{c}^{k*-1} \rangle = \langle \hat{c}^k, \partial_k^* \hat{c}^{k*-1} \rangle = \langle \hat{c}^k, \delta_{k-1} \hat{c}^{k*-1} \rangle. \tag{A8}$$

The coboundary operator δ_k has a fundamental property similar to the boundary operator ∂_k

$$\delta_k \circ \delta_{k-1}(\cdot) = 0. \tag{A9}$$

k-cochains whose coboundaries are zero are called *k-cocycles*.

k-cochains obtained from $(k-1)$-cochains via the map δ_{k-1} are referred to as *k-coboundaries*. As an illustration of the map $\delta_0 \hat{c}^{0*} \to \hat{c}^{1*}$, consider an arbitrary vertex a and define an arc as *incident from* a if the vertex a is the initial and not the terminal vertex, while *incident to* a implies that a is the terminal and not the initial vertex. Furthermore, if we denote $\omega^+(A)$ as the set of arcs incident from a set of vertices A and $\omega^-(A)$ as the set of arcs incident to A, then the 1-coboundary is a set of arcs of the form $\omega(A)$ which is nonempty and can be partitioned into two classes, $\omega^+(A)$ and $\omega^-(A)$. With every coboundary of m arcs one may associate a vector $\omega = (\omega_1, \ldots, \omega_m)$ such that

$$\omega_i = \begin{pmatrix} 0 & \text{if } i \notin \omega(A) \\ 1 & \text{if } i \in \omega^+(A) \\ -1 & \text{if } i \in \omega^-(A) \end{pmatrix}. \tag{A10}$$

The ω_i are *elementary coboundaries* and have the property $\omega_i \cap \omega_j = \phi$, $i \neq j$.

Theorem A.2: Every 1-coboundary is a direct sum of elementary coboundaries.

With this background it can easily be shown that the following theorems hold true.

Theorems A.3: Let $Y^1 = \ker(\partial_1)$ be the cycle subspace of V^1 and $D^{1*} = \text{Image}(\delta_0)$ be the coboundary subspace of V^{1*}. Then for a connected linear graph Γ with m arcs and n nodes:

(1) The inner product of a 1-cycle and a 1-coboundary is 0, i.e., $\langle \hat{y}^1, \hat{d}^{1*} \rangle = 0$, $\forall \hat{y}^1 \in Y^1$, $d^{1*} \in D^{1*}$.

(2) Any 1-chain orthogonal to every 1-coboundary is a 1-cycle, i.e., if $\langle \hat{c}^1, \partial^{1*} \rangle = 0$, $\forall \hat{d}^{1*} \in D^{1*}$, then $\hat{c}^1 \in Y^1$.

(3) Any 1-cochain orthogonal to every 1-cycle is a 1-coboundary, i.e., if $\langle \hat{c}^{1*}, \hat{y}^1 \rangle = 0$, $\forall \hat{y}^1 \in Y^1$, then $\hat{c}^{1*} \in D^{1*}$.

(4) Every 1-chain is uniquely expressible as a direct sum of a 1-cycle and a 1-coboundary.

(5) The dimensions of the cycle and coboundary spaces Y^1 and D^{1*}, respectively, are

$$\dim Y^1 = m - n + 1, \quad \dim D^1 = n - 1.$$

APPENDIX B

The time independent Schrödinger equation for an interacting N-particle system can be derived from the Lagrangian (3.14)

$$L = \frac{\hbar^2}{2m} \sum_{i=1}^N \left(\frac{\partial \psi(\mathbf{x}_1, \ldots, \mathbf{x}_N)}{\partial \mathbf{x}_i} \right)^2$$

$$+ [V(\mathbf{x}_1, \ldots, \mathbf{x}_N) - E] \psi^2(\mathbf{x}_1, \ldots, \mathbf{x}_N), \quad \mathbf{X}_i \in R^3 \tag{B1}$$

by using the method of Lagrangian multipliers. By writing

$$\frac{\partial \psi(\mathbf{x}_1, \ldots, \mathbf{x}_N)}{\partial \mathbf{x}_i} = W_i \tag{B2}$$

the original Lagrangian may be modified to

$$L' = L + \sum_{i=1}^N P_i \frac{(\partial \psi(\mathbf{x}_1, \ldots, \mathbf{x}_N) - W_i)}{\partial \mathbf{x}_i}$$

$$= \frac{\hbar^2}{2m} \sum_{i=1}^N W_i^2 + [V(\mathbf{x}_1, \ldots, \mathbf{x}_N) - E] \psi^2(\mathbf{x}_1, \ldots, \mathbf{x}_N) \tag{B3}$$

$$+ \sum_{i=1}^N P_i \frac{(\partial \psi(\mathbf{x}_1, \ldots, \mathbf{x}_N) - W_i)}{\partial \mathbf{x}_i},$$

the P_i acting as the Lagrangian multipliers. Since the W_i appear only as algebraic variables, without derivatives, they are kinosthenic and

$$\frac{\partial L'}{\partial W_i} = \frac{\hbar^2 W_i}{m} - P_i = 0. \tag{B4a}$$

Thus,

$$W_i = m P_i / \hbar^2, \tag{B4b}$$

and (B3) becomes

$$L' = \frac{m}{2\hbar^2} \sum_{i=1}^{N} P_i^2 + (V(\mathbf{x}_1, \ldots, \mathbf{x}_N) - E)\psi^2(\mathbf{x}_1, \ldots, \mathbf{x}_N)$$

$$- \frac{m}{\hbar^2} \sum_{i=1}^{N} P_i^2 + \sum_{i=1}^{N} P_i \frac{\partial \psi}{\partial \mathbf{x}_i}(\mathbf{x}_1, \ldots, \mathbf{x}_N)$$

$$= - \frac{m}{2\hbar^2} \sum_{i=1}^{N} P_i^2 + [V(\mathbf{x}_1, \ldots, \mathbf{x}_N) - E]\psi^2(\mathbf{x}_1, \ldots, \mathbf{x}_N)$$

$$+ \sum_{i=1}^{N} P_i \frac{\partial \psi}{\partial \mathbf{x}_i}(\mathbf{x}_1, \ldots, \mathbf{x}_N)$$

$$= H(P_i, \psi) + \sum_{i=1}^{N} P_i \frac{\partial \psi}{\partial \mathbf{x}_i}(\mathbf{x}_1, \ldots, \mathbf{x}_N).$$

Variation with respect to P_i yields

$$- \frac{m}{\hbar^2} P_i + \frac{\partial \psi}{\partial \mathbf{x}_i}(\mathbf{x}_1, \ldots, \mathbf{x}_N) = 0 \qquad \text{(B6a)}$$

and variation with respect to $\psi(\mathbf{x}_1, \ldots, \mathbf{x}_N)$ yields

$$2[V(\mathbf{x}_1, \ldots, \mathbf{x}_N) - E]\psi(\mathbf{x}_1, \ldots, \mathbf{x}_N) - \sum_{i=1}^{N} \frac{\partial P_i}{\partial \mathbf{x}_i} = 0. \quad \text{(B6b)}$$

By substitution of (B6a) into (B6b) we have Q.E.D.

[1]G. Kron, Phys. Rev. **64**, 126 (1943); Proc. Inst. Radio Engrs. Aust. **32**, 289 (1944); J. Appl. Mech. **11**, 149 (1944); Phys. Rev. **67**, 39 (1945); J. Aeronaut. Sci. **12**, 221 (1945); J. Chem. Phys. **14**, 19 (1946); Electron. Eng. **67**, 672 (1948).

[2]A.G.J. MacFarlane, *Dynamical System Models* (Harrap, London, 1970), p. 278.

[3]J.R. Platt, *Free Electron Theory of Conjugated Molecules* (Wiley, New York, 1964).

[4]C.A. Coulson, Proc. Phys. Soc. (London) **67**, 608 (1954); **68**, 1129 (1953).

[5]G. Della Riccia, Proceedings of the Semiconductor Conference, Exeter (Inst. of Physics and the Phys. Soc., London, 1962), p. 570.

[6]E.W. Montroll, J. Math. Phys. **11**, 635 (1970).

[7]E.W. Montroll, J. Phys. Chem. Solids **34**, 597 (1973).

[8]A.B. Budgor, J. Math. Phys. **16**, 2153 (1975).

[9]P. Penfield, Jr., R. Spence, and S. Duinker, *Tellegen's Theorem and Electrical Networks*, Research Monograph #58 (MIT Press, Cambridge, Mass., 1970).

[10]J.P. Roth, Proc. Natl. Acad. Sci. USA **41**, 518 (1955).

[11]F. Branin, in *Proceedings of the Symposium on Generalized Networks, Brooklyn, N.Y.* (Polytechnic Press of the Polytechnic Inst. of Brooklyn, New York, 1966).

[12]S. Smale, J. Diff. Geom **7**, 193 (1972).

[13]R.K. Brayton and J.K. Moser, Quart. Appl. Math. **22**, 1, 81 (1964).

[14]C. Berge and A. Ghouila-Houri, *Programming, Games, and Transportation Networks* (Methuen, London, 1965).

[15]C. Lanczos, *The Variational Principles of Mechanics* (University of Toronto Press, Toronto, 1970).

[16]A.A. Abrikosov, L.P. Gorkov, and I.E. Dzyaloshinski, *Quantum Field Theoretical Methods in Statistical Physics* (Pergamon, New York, 1965).

Bogolubov–Parasiuk theorem in the α-parametric representation

M. C. Bergere*† and Yuk-Ming P. Lam‡§

Service de Physique Theorique, Centre d'Etudes Nucléaires de Saclay, BP n 2-91190 Gif-sur-Yvette, France
(Received 6 October 1975)

A renormalized Feynman amplitude expressed in the α-parameters is defined by introducing a subtraction operator acting directly upon the α-integrand. Different forms of this subtraction operator are discussed. We define the isotropic and nonisotropic normal products and we give a more general oversubtraction rule which ensures both the absolute convergence of the amplitude and the Bogolubov, Parasiuk and Hepp recurrence. The proof of absolute convergence of the amplitude is performed using Hepp's sectors and equivalence classes of nests.

I. INTRODUCTION

The Bogolubov-Parasiuk[1] theorem on renormalization, which tells how to extract a finite part from a divergent Feynman amplitude in such a way that the resulting field theory satisfies Poincaré invariance, unitarity, and causality, was first established in 1957, and completed later by Hepp[2] in 1966. In their proof, they first regularize with a cutoff the divergent Feynman amplitude, and then, they recursively subtract away a Taylor expansion in the external momenta of the graph and its subgraphs.

In 1973, Zimmermann[3] gave an explicit solution to the above recursive process in the form of Taylor subtractions of the integrand expressed in momentum space. The introduction of a cutoff is then completely avoided since the resulting integral is proved to be absolutely convergent.

An independent way of renormalizing a divergent Feynman amplitude by performing subtractions on the integrand, not according to Taylor expansion in the momenta but according to Laurent expansion in the Schwinger α parameters was given in Ref. 4. This scheme is a generalization of the work of Appelquist[5] completed by Anikin, Polivanov, and Zavialov.[5] The introduction of generalized Taylor operators[4] acting upon the integrand expressed in the α-Schwinger parameters turned out to be very convenient to solve various problems related to the technique of renormalization (for instance, the asymptotic behavior of the Feynman amplitudes when the scaling parameter tends to infinity[6,7]).

In this paper, we intend to give a simple proof of the absolute convergence of the renormalized Feynman amplitude expressed in the α parameters. This proof is generalized to the case of oversubtractions (isotropic[3] and nonisotropic normal products[8]) as well as to the case of soft mass subtractions.[9]

In Sec. II, we remind to the reader, the power counting theorem for convergent Feynman amplitudes expressed in the α parameters. The renormalized Feynman amplitude is defined in Sec. III, and the proof of absolute convergence is given. We restrict ourself to Euclidean space; the generalization to Minkowsky space in the sense of distribution was performed by Hepp, and is not exposed in this paper. In Sec. IV, we

discuss an oversubtraction rule which preserves both, the absolute convergence of the amplitude and the recurrence of Bogolubov, Parasiuk, and Hepp (BPH) in terms of generalized vertices. Finally, in Sec. V, we prove that the renormalized amplitude as defined in Sec. III satisfies BPH recurrence. Two appendices are devoted to technical problems.

II. THEOREM ON THE ABSOLUTE CONVERGENCE OF THE MULTIPLE INTEGRAL

$$\int_0^\infty \prod_{a=1}^l d\alpha_a \exp(-\sum_{a=1}^l \alpha_a m_a^2) Z(\alpha).$$

The above integral is performed over a domain defined by

$$0 \leq \alpha_i < \infty \quad \text{for } i=1,\ldots,l, \tag{2.1}$$

Although the following considerations apply to general functions of the above kind, it is convenient to keep the usual Feynman diagram terminology. Let us call graph the set $G = \{1,\ldots,l\}$.

Definition II 1: A subset S of integers between 1 and l inclusive is called a subgraph. $l(S)$ is defined to be the number of integers in S.

Definition II 2: A nonempty collection of subgraphs $S_1,\ldots,S_t$ is called a nest if $i \geq j$ implies $S_i \supseteq S_j$.

Definition II 3: A function $Z(\alpha)$ is said to have the "Taylor series property" with respect to a subgraph S if there exists a complex number $\mu(S)$ such that

$$\rho^{-\mu(S)} Z(\alpha)\Big|_{\alpha_a \to \rho^2 \alpha_a \text{ for } a \in S} \tag{2.2}$$

has a Taylor series in ρ at $\rho = 0$ and does not vanish at $\rho = 0$.

Definition II 4: A function $Z(\alpha)$ is said to have the "simultaneous Taylor series property" with respect to a nest N, if there exists complex numbers $\mu(S)$ for every subgraph $S \in N$ such that

$$\prod_{S \in N} \rho_\varphi^{-\mu(S)} Z(\alpha)\Big|_{\alpha_a \to \rho_S^2 \alpha_a \text{ for } a \in S} \tag{2.3}$$

has a simultaneous Taylor series in the variables ρ_S at $\rho_S = 0$ and does not vanish when all $\rho_S = 0$.

Example: The function $(\alpha_1 + \alpha_2)^{-1}$ has the "simulta-

neous Taylor series property" with respect to the nest consisting of subgraphs $S_1 = \{1\}$ and $S_2 = \{1,2\}$.

Theorem: If (i) $Z(\alpha)$ exists for $0 < \alpha_i < \infty$; (ii) $Z(\alpha)$ is polynomially bounded when arbitrary subsets of $\{\alpha_1, \ldots, \alpha_i\}$ are scaled to ∞; (iii) $Z(\alpha)$ has the "simultaneous Taylor series property" with respect to every nest N of subgraphs; (iv) the superficial degree of divergence

$$\omega(S) = -\mu(S) - 2l(S) \tag{2.4}$$

is strictly negative. Then, the integral

$$I = \int_0^\infty \prod_{a=1}^l d\alpha_a \exp(-\sum_{a=1}^l \alpha_a m_a^2) Z(\alpha)$$

is absolutely convergent provided $m_a \neq 0$ for all a. This theorem is the so-called power-counting theorem.

Proof: We decompose the domain of integration into Hepp's sectors,[2] each defined by a permutation $\{i_1, i_2, \ldots, i_l\}$ of the integers $\{1, 2, \ldots, l\}$:

$$S_\ell = \{\alpha \mid 0 \leq \alpha_{i_1} \leq \cdots \leq \alpha_{i_l}\}, \tag{2.5}$$

and we perform in each sector the change of variables

$$\alpha_{i_j} = \prod_{k=j}^l \beta_k^2,$$

$$d\alpha_{i_j} = 2\beta_j \, d\beta_j \prod_{k=j+1}^l \beta_k^2. \tag{2.6}$$

For each sector the integration domain becomes

$$0 \leq \beta_l < \infty,$$
$$0 \leq \beta_j \leq 1 \quad \text{for } j < l, \tag{2.7}$$

and the Jacobian turns out to be $2^l \prod_{j=1}^l \beta_j^{2l-1}$. In a given sector, the change of variables transform $Z(\alpha)$ into

$$\prod_{j=1}^l \beta_j^{\mu(R^j)} Z'(\beta_i), \tag{2.8}$$

where the subgraphs $R^j = \{i_1, \ldots, i_j\}$, and where the function $Z'(\beta_i)$ exists in the domain (2.7) and is polynomially bounded when $\beta_l \to \infty$.

Then

$$I = \sum_\ell I_\ell, \tag{2.9}$$

with

$$I_\ell = 2^l \int_0^\infty d\beta_l \int_0^1 \prod_{j=1}^{l-1} d\beta_j \prod_{j=1}^l \beta_j^{\mu(R^j)+2j-1}$$

$$\times \exp[-\beta_l^2(m_{i_l}^2 + \sum_{j=1}^{l-1} \beta_j^2 \cdots \beta_{l-1}^2 m_{i_j}^2)] Z'(\beta_i). \tag{2.10}$$

The absolute convergence of I_ℓ is now straightforward. Note that the property that $Z'(\beta_i)$ exists when the β's equal zero is due to the "simultaneous Taylor series property." In (2.10), $\mu(R^j) + 2j$ equals $\mu(R^j) + 2l(R^j) = -\omega(R^j) > 0$.

III. R OPERATOR

This section is devoted to the definition of a finite part of the integral

$$I = \int_0^\infty \prod_{a=1}^l d\alpha_a \exp(-\sum_{a=1}^l \alpha_a m_a^2) Z(\alpha),$$

where at least one $\omega(S)$ is positive or null and where all masses differ from zero. It is the purpose of Secs. IV and V to show that the finite part chosen here (among others) is related in perturbative quantum field theory to the renormalization process.

Definition III 1: If $f(x)$ is such that $x^{-\mu} f(x)$ has a Taylor series at $x = 0$, then for any integer n, we define the generalized Taylor operator τ_x^n on $f(x)$ by

$$\tau_x^n f(x) = x^\mu T_x^{n-E'(\mu)} (x^{-\mu} f(x)), \tag{3.1}$$

where $E'(\mu)$ is the smallest integer larger or equal to $\mathrm{Re}\,\mu$ and $T^{n-E'(\mu)}(x^{-\mu} f(x))$ are the first $n - E'(\mu) + 1$ term of the Taylor expansion of $x^{-\mu} f(x)$ at $x = 0$.

Definition III 2: If a function $Z(\alpha)$ has "the Taylor series property" with respect to a subgraph S, we define for any integer n, τ_S^n on $Z(\alpha)$ by

$$\tau_S^n Z(\alpha) = \{\tau_\rho^n Z(\alpha) \mid_{\alpha_a \to \rho^2 \alpha_a, \ a \in S}\}_{\rho=1}. \tag{3.2}$$

Further properties of the generalized Taylor operators τ can be found in Appendix A.

Theorem: If (i) $Z(\alpha)$ is infinitely differentiable for $0 < \alpha_i < \infty$; (ii) $Z(\alpha)$ and its α derivatives are polynomially bounded when arbitrary subsets of $\{\alpha_1, \ldots, \alpha_i\}$ are scaled to ∞; (iii) $Z(\alpha)$ has the "simultaneous Taylor series property" with respect to every nest N of subgraphs; (iv) for every subgraph S, $C(S)$ is a non-negative integer satisfying $S \supset S' \to C(S) \geq C(S')$, then, the integral

$$I_R = \int_0^\infty \prod_{a=1}^l d\alpha_a \exp(-\sum_{a=1}^l \alpha_a m_a^2) R\{Z(\alpha)\}$$

is absolutely convergent provided $m_a \neq 0$ for all a.

In the above integral, R is the subtraction operator which we define as

$$R = (1 + \sum_N \prod_{S \in N} (-\tau_S^{-2l(S)+c(S)})),$$

where we sum over all possible nests of subgraphs S; $C(S)$ takes care of possible oversubtractions.

We show in Appendix B that the operator R can also be be defined as

$$R = \prod_{S \in G} (1 - \tau_S^{-2l(S)+c(S)}),$$

where the product runs over all the $(2^l - 1)$ subgraphs of G, provided that condition (iv) is replaced by the stronger condition $C(S_1 \cup S_2) \geq C(S_1) + C(S_2) - C(S_1 \cap S_2)$ and $C(S) \geq 0$. In that case, the product which defines R can be taken in any order.

The plan of the proof of absolute convergence is as follows: (a) decomposition of the α-integration domain into Hepp's sectors (2-5), (2-6), and definition of the nested subgraphs $R^j = \{i_1, \ldots, i_j\}$ for $j = 1, \ldots, l$; (b) construction of a maximal nest G, from a given nest N and a given sector; (c) definition of equivalence classes of nests N with the same maximal nest G, and summation over all nests of the same equivalence class; (d) application of the τ operators upon $Z(\alpha)$ and proof

of absolute convergence in each sector for each equivalence class. The first step has already been accomplished in Sec. II; let us proceed with the second step.

A. Ω construction

Given any three subgraphs R, S, and T, we define the subgraph

$$\omega^R(S, T) = S \cap (R \cup T),\qquad (3.3)$$

and we have the property

$$\omega^R(S, \omega^R(S, T)) = \omega^R(\omega^R(S, T), T) = \omega^R(S, T),\qquad (3.4)$$

If moreover $S \supseteq T$, then

$$\omega^R(S, T) = S \cap (R \cup T) = T \cup (R \cap S),\qquad (3.5)$$

and we have

$$T \subseteq \omega^R(S, T) \subseteq S,\qquad (3.6a)$$

$$\omega^R(S, T) = S \longrightarrow S \cup R = T \cup R,\qquad (3.6b)$$

$$\omega^R(S, T) = T \longrightarrow S \cap R = T \cap R.\qquad (3.6c)$$

Definition III 3: For every nest $N = \{T_1, \ldots, T_l\}$, we define the nest $N' = \{T_0, T_1, \ldots, T_l, T_{l+1}\}$, where T_0 is the empty subgraph Φ and $T_{l+1} = G' = \{1, 2, \ldots, l, l+1\} \supset G$. Given a nest $R = (R^0 = \Phi, R^1, \ldots, R^r, R^{r+1} = G')$ and a nest N', we define a R-maximal nest G as

$$G = \Omega^R(N') = \{\omega^{R^i}(T_j, T_{j-1}): i = 0, \ldots, r+1; j = 1, \ldots, t+1\}.\qquad (3.7)$$

The presence of R^0 and R^{r+1} in the nest R is such that

$$N' \subseteq \Omega^R(N').\qquad (3.8)$$

On the other hand, (3.6a) together with

$$\omega^{R^i}(T_j, T_{j-1}) \subseteq \omega^{R^{i'}}(T_j, T_{j-1}) \text{ for } i < i',\qquad (3.9)$$

makes of $\Omega^R(N')$ a nest. Now G is a R-maximal nest, that is by definition

$$\Omega^R(G) = G.\qquad (3.10)$$

Indeed, given two consecutive elements of the nest G which we call $\omega^{R^i}(T_j, T_{j-1})$ and $\omega^{R^{i+1}}(T_j, T_{j-1})$, it can be shown using (3.5) that

$$\omega^{R^k}(\omega^{R^{i+1}}(T_j, T_{j-1}), \omega^{R^i}(T_j, T_{j-1}))$$

$$= \begin{cases} \omega^{R^i}(T_j, T_{j-1}) & \text{for } k \leqslant i, \\ \omega^{R^{i+1}}(T_j, T_{j-1}) & \text{for } k \geqslant i+1. \end{cases}\qquad (3.11)$$

Next, we may partition G into three parts in the following way.

Let us now rename the distinct elements of G as $G = \{T_0 = \Phi, T_1, \ldots, T_t, T_{t+1} = G'\}$ and since G is R maximal, we have

$$\omega^{R^i}(T_j, T_{j-1}) = T_j \text{ or } T_{j-1} \text{ for } j = 1, \ldots, t,$$

$$\omega^{R^i}(T_{t+1}, T_t) = T_t \text{ for } i < r+1.\qquad (3.12)$$

Then, because of (3.9) there exists for each $j = 1, \ldots, t+1$ a number $p(j)$ such that $0 \leqslant p(j) \leqslant r+1$, and

$$\omega^{R^i}(T_j, T_{j-1}) = \begin{cases} T_{j-1} & \text{for } i \leqslant p(j), \\ T_j & \text{for } i > p(j), \end{cases}\qquad (3.13)$$

where (3.13) is nothing but a rephrasing of the property mentioned in (3.11). Note that $p(t+1) = r$.

Letting $p(0) = r+1$ and $p(t+2) = 0$, we define

$$K = \{T_j: j = 0, \ldots, t+1; p(j) > p(j+1)\},\qquad (3.14)$$

$$H = \{T_j: j = 1, \ldots, t; p(j) < p(j+1)\},\qquad (3.15)$$

$$B = G - K - H.\qquad (3.16)$$

Note that $T_{t+1} = G' \in K$ always; $T_0 = \Phi \in K$ also since $p(1) \leqslant r$ because $i = r+1$ is such that $\omega^{R^i}(T_1, \Phi = T_0) = T_1$.

Let us now state some properties of the subnests K and H. We define

$$K^i = \{T_j: j = 1, \ldots, t; \omega^{R^i}(T_j, T_{j-1})$$

$$= T_{j-1}; \omega^{R^i}(T_{j+1}, T_j) = T_{j+1}\}$$

$$\cup \left\{ \begin{array}{ll} \{T_0\} & \text{if } \omega^{R^i}(T_1, T_0) = T_1 \\ 0 & \text{otherwise} \end{array} \right\}$$

$$\cup \left\{ \begin{array}{ll} \{T_{t+1}\} & \text{if } \omega^{R^i}(T_{t+1}, T_t) = T_t \\ 0 & \text{otherwise} \end{array} \right\}.\qquad (3.17)$$

Then, $K^0 = \{G'\}$, $K^{r+1} = \{\Phi\}$, $G' \in K^i$ for $i \neq r+1$ and

$$K = \bigcup_{i=0}^{r+1} K^i.\qquad (3.18)$$

One element T_j can belong to several K^i; we label the elements of K^i by

$$K^i = \{K_1^i, \ldots, K_{r_i}^i = G'\} \text{ for } i = 1, \ldots, r.\qquad (3.19)$$

Similarly, we define

$$H^i = \{T_j: j = 1, \ldots, t; \omega^{R^i}(T_j, T_{j-1}) = \omega^{R^i}(T_{j+1}, T_j) = T_j\},\qquad (3.20)$$

such that H^0 and H^{r+1} are empty and

$$H = \bigcup_{i=1}^{r} H^i.\qquad (3.21)$$

It can be proved[10] for $1 \leqslant i \leqslant r$ that in between two consecutive elements of K^i is one and only one element of H^i; in between two consecutive elements of H^i is one and only one element of K^i; K^i has at least two elements; H^i is never empty since there is one element less in H^i than in K^i; and $\omega^{R^i}(K_{j+1}^i, K_j^i) \in H^i$ for $j = 1, \ldots, r_i - 1$. We define

$$H_j^i = \omega^{R^i}(K_{j+1}^i, K_j^i) \text{ for } 1 \leqslant i \leqslant r \text{ and } 1 \leqslant j \leqslant r_i - 1.\qquad (3.22)$$

Again, several H_j^i can represent the same subgraphs.

Two nests N_1' and N_2' are said to be R equivalent if $\Omega^R(N_1') = \Omega^R(N_2')$. This equivalence relation partitions the set W of all nests into a set W/R of R-equivalence classes. Each equivalence class Γ is characterized by the R-maximal nest $G = \Omega^R(N) \forall N \in \Gamma$. The largest and smallest nests in this R-equivalence class Γ are, respectively, G and $B \cup K$. Then if $N' \in \Gamma$,

$$B \cup K \subseteq N' \subseteq G,\qquad (3.23)$$

and conversely if H' is any subset of H, $N' = B \cup K \cup H' \in \Gamma$. To resume the development of the Ω construction, let us now state the following lemmas:

Lemma II 1: The R-equivalence class Γ of nests N'

is isomorphic to the set of all subsets of H in the sense that N' in Γ corresponds to that H' in H satisfying $N' = B \cup K \cup H'$.

Lemma II 2:

$$\sum_{j=1}^{r_i-1} [l(H_j^i) - l(K_j^i)] = l(R^i) \text{ for } i = 1, \ldots, r. \quad (3.24)$$

The proof of lemma II 2 follows the definition (3.22) and the topologic relation $l(S_1 \cup S_2) = l(S_1) + l(S_2) - l(S_1 \cap S_2)$ for any subgraphs S_1 and S_2.

Given a sector g as defined in (2.5) we define the nest

$$R(g) = (R^0 = \phi, R^1, \ldots, R^i = G, R^{i+1} = G'),$$
$$R^j = \{i_1, \ldots, i_j\}. \quad (3.25)$$

For each $R(g)$-equivalent class $\Gamma \in W/R(g)$, we consider the space consisting of coordinates β_i for each $R^i \in \{R(g) - \{\Phi\} - \{G'\}\}$, ξ_B for each $B \in \mathcal{B}$, σ_j^i for each $K_j^i \in \mathcal{K}$ and χ_j^i for each $H_j^i \in \mathcal{H}$ ($i = 1, \ldots, l; j = 1, \ldots, r_i - 1$). We define the transformation $(g\Gamma)$ from the $(\alpha_1, \ldots, \alpha_l)$-space to the above space by

$$\alpha_a = \begin{pmatrix} \prod_{\substack{i=1,\ldots,l \\ a \in R^i}} \beta_i & \prod_{\substack{B \in \mathcal{B} \\ a \in}} \xi_B & \left\{ \prod_{\substack{i=1,\ldots,l \\ j=1,\ldots,r_i-1 \\ a \in K_j^i}} \sigma_j^i \right\} & \left\{ \prod_{\substack{i=1,\ldots,l \\ j=1,\ldots,r_i-1 \\ a \in H_j^i}} \chi_j^i \right\} \end{pmatrix}^2. \quad (3.26)$$

Lemma II 3: Under the transformation $(g\Gamma)$, the function $Z(\alpha)$ is transformed to a function of the variables ξ_B, and of the group of variables σ_j^i/β_i and $\beta_i\chi_j^i$ ($i = 1, \ldots, l; j = 1, \ldots, r_i - 1$) that is

$$Z(\alpha) \overset{(g\Gamma)}{=} Z^{g\Gamma}(\xi_B, \sigma_j^i/\beta_i, \beta_i\chi_j^i). \quad (3.27)$$

Proof: We give ourselves a line a. For a given R^i, we call K_j^i the smallest element of K^i which contains the line a. Then from (3.22) and the nest property, a is also in H_j^i, K_{j+1}^i, $H_{j+1}^i \cdots K_{r_i-1}^i$, $H_{r_i-1}^i$. The problem is to know whether a is in H_{j-1}^i. From (3.22) again, with j replaced by $j - 1$, if a is in R^i, then a is in H_{j-1}^i; if a is not in R^i, a is not in H_{j-1}^i. Consequently,

$$Z(\alpha) \overset{(g\Gamma)^1}{=} Z^{g\Gamma}(\xi_B \prod_{p=j}^{r_i-1} \sigma_p^i\chi_p^i, \beta_i\chi_{j-1}^i) \text{ if } a \text{ is in } R^i$$

and

$$Z(\alpha) \overset{(g\Gamma)}{=} Z^{g\Gamma}(\xi_B, \prod_{p=j}^{r_i-1} \sigma_p^i\chi_p^i) \text{ if } a \text{ is not in } R^i.$$

Both forms can be rewritten under the form (3.27).

B. "Simultaneous Taylor series property"

If $Z(\alpha)$ has the "simultaneous Taylor series property" with respect to a $R(g)$ equivalence class $\Gamma \in W/R(g)$, that is with respect to its $R(g)$ maximal nest $\mathcal{G}$, then

$$Z^{g\Gamma}(\xi_B, \sigma_j^i/\beta_i, \beta_i\chi_j^i) = \prod_{B \in \mathcal{B}} (\xi_B)^{\mu(B)}$$

$$\times \prod_{i=1}^{l} \prod_{j=1}^{r_i-1} \left[\left(\frac{\sigma_j^i}{\beta_i} \right)^{\mu(K_j^i)} (\beta_i\chi_j^i)^{\mu(H_j^i)} \right]$$

$$\times \sum_{d_B=0}^{\infty} \sum_{a_j^i=0}^{\infty} (\xi_B)^{d_B} \left(\frac{\sigma_j^i}{\beta_i} \right)^{a_j^i} \Lambda_{\{d,a\}}^{g\Gamma}(\beta_i\chi_j^i), \quad (3.28)$$

where $\mu(S)$ has been defined in (2.2) and where, of course,

$$\Lambda_{\{d,a\}}^{g\Gamma}(\beta_i\chi_j^i) = \prod_{B \in \mathcal{B}} \frac{1}{d_B!} \left(\frac{\partial}{\partial \xi_B} \right)^{d_B} \prod_{i=1}^{l} \prod_{j=1}^{r_i-1} \frac{1}{a_j^i!}$$

$$\times \beta_i \frac{\partial}{\partial \sigma_j} \left\{ \prod_{B \in \mathcal{B}} \xi_B^{-\mu(B)} \prod_{i=1}^{l} \prod_{j=1}^{r_i-1} \left[\left(\frac{\sigma_j^i}{\beta_i} \right)^{-\mu(K_j^i)} \right. \right.$$

$$\times (\beta_i\chi_j^i)^{-\mu(H_j^i)} \left] Z^{g\Gamma} \left(\xi_B, \frac{\sigma_j^i}{\beta_i}, \beta_i\chi_j^i \right) \right\} \bigg|_{\xi_B = \sigma_j^i = 0} \quad (3.29)$$

$\Lambda_{\{d,a\}}^{g\Gamma}(\beta_i\chi_j^i)$ has a simultaneous Taylor expansion in $\beta_i\chi_j^i$.

For a given sector g, let us call (g) the transformation defined in (2.6).

Lemma II 4: If $Z(\alpha)$ has the "simultaneous Taylor series property" with respect to a $R(g)$-equivalence class $\Gamma \in W/R(g)$, and if $C(S)$ is an integer for every $S \in N' \in \Gamma$, then

$$\sum_{N' \in \Gamma} \prod_{S \in N} \left(-\tau_S^{-2l(S)+C(S)} \right) Z(\alpha) \bigg|_{\alpha \overset{(g)}{=} \beta}$$

$$= \sum_{a_j^i=0}^{m(K_j^i)} \sum_{d_B=0}^{m(B)} \prod_{i=1}^{l} \beta_i^{p_i} \int_0^1 \prod_{\substack{i=1,\ldots,l \\ j=1,\ldots,r_i-1 \\ m(H_j^i) \geq 0}} \left[d\chi_j^i \frac{(1-\chi_j^i)^{m(H_j^i)}}{m(H_j^i)!} \left(\frac{\partial}{\partial \beta_i\chi_j^i} \right)^{m(H_j^i)+1} \right] \Lambda_{\{d,a\}}^{\Gamma}(\beta^i\chi_j^i) \bigg|_{\chi_j^i=1 \text{ for } m(H_j^i) < 0} \quad (3.30)$$

if all $m(K_j^i)$ and $m(B)$ are positive or null and zero otherwise. In (3.30),

$$m(S) = -2l(S) + C(S) - E'[\mu(S)], \quad (3.31)$$

and for $i = 1, \ldots, l$,

$$p_i = \sum_{j=1}^{r_i-1} [\mu(H_j^i) - \mu(K_j^i) - a_j^i] + \sum_{\substack{j=1,\ldots,r_i-1 \\ m(H_j^i) \geq 0}} [m(H_j^i) + 1]. \quad (3.32)$$

Proof: The proof of this lemma is simple; the τ operators relative to subgraphs of a nest commute. From lemma II 1, we can write

$$\sum_{N' \in \Gamma} \prod_{S \in N} (-\tau_S) = \prod_{S \in \mathcal{B} \cup \mathcal{K}} (-\tau_S) \prod_{i=1}^{l} \prod_{j=1}^{r_i-1} (1 - \tau_{H_j^i}). \quad (3.33)$$

Then from (3.28) and the remainder theorem for Taylor expansions, we obtain (3.30).

Lemma II 5: Under the assumption of lemma II 4 and further assuming that $C(\mathcal{S})$ is a non-negative integer such that $\mathcal{S} \supset \mathcal{S}'$ implies $C(\mathcal{S}) \geqslant C(\mathcal{S}')$, then

$$p_i > -2i. \tag{3.34}$$

Proof: From $a_j^i \leqslant m(K_j^i)$ and from $m(H_j^i) + 1 \leqslant 0$ for $m(H_j^i) < 0$, we obtain

$$p_i \geqslant \sum_{j=1}^{r_i-1} [\mu(H_j^i) - \mu(K_j^i) + 2l(K_j^i) - C(K_j^i) + E'[\mu(K_j^i)] - 2l(H_j^i) + C(H_j^i) - E'[\mu(H_j^i)] + 1].$$

Then, using $\nu \leqslant E'(\nu) < 1 + \nu$,

$$p_i > \sum_{j=1}^{r_i-1} [2l(K_j^i) - 2l(H_j^i) + C(H_j^i) - C(K_j^i)].$$

From lemma II 2 and from (3.22), which tells that $H_j^i \supseteq K_j^i$, that is, $C(H_j^i) \geqslant C(K_j^i)$ for $j \neq 1$ and for $j = 1$ if $K_1^i \neq \Phi$ [if $K_1^i = \Phi$, $C(K_1^i) = 0$, and $C(H_1^i) \geqslant 0$], we prove (3.34).

C. Proof of the theorem

Given a sector $\mathcal{S}_g$ of the integration domain as defined in (2.5), we consider a $R(g)$-equivalent class Γ of subgraphs.

Then, using lemma II 4 and the Jacobian $2^l \Pi_{i=1}^l \beta_i^{2i-1}$ of the transformation $\alpha \overset{(g)}{\rightleftharpoons} \beta$, we get

$$\int_{\mathcal{S}_g} \prod_{a=1}^l d\alpha_a \exp(-\sum_{a=1}^l \alpha_a m_a^2) \sum_{N' \in \Gamma \mathcal{S}} \prod_{\in N} (-\tau^{-2l(\mathcal{S}) + c(\mathcal{S})}) Z(\alpha) = 2 \sum_{a_j^i = 0}^{m(K_j^i)} \sum_{d_B = 0}^{m(B)} \int_0^\infty d\beta_i \, \beta_i^{p_i + 2l - 1}$$

$$\times \int_0^1 \prod_{i=1}^{l-1} (d\beta_i \beta_i^{p_i + 2i - 1}) \exp[-\beta_i^2(m_{i_i}^2 + \sum_{j=1}^{l-1} \beta_j^2 \cdots \beta_{l-1}^2 m_{ij}^2)]$$

$$\times \int_0^1 \prod_{\substack{i,j \\ m(H_j^i) \geqslant 0}} \left[d\chi_j^i \frac{(1 - \chi_j^i)^{m(H_j^i)}}{m(H_j^i)!} \left(\frac{\partial}{\partial\beta_i\chi_j^i}\right)^{m(H_j^i)+1} \right] \Lambda_{\{d,a\}}^{g\Gamma}(\beta_i'\chi_{j'}^{i'}) \Big|_{\chi_{j'}^{i'} = 1 \ \text{for} \ m(H_{j'}^{i'}) < 0}. \tag{3.35}$$

Assumptions i and ii of the theorem imply

$$\left| \prod_{\substack{i,j \\ m(H_j^i) \geqslant 0}} \left(\frac{\partial}{\partial\beta_i\chi_j^i}\right)^{m(H_j^i)+1} \Lambda_{\{d,a\}}^{g\Gamma}(\beta_i, \chi_{j'}^{i'}) \right| < P(\beta_i) \quad \text{for} \ \beta\chi \neq 0, \tag{3.36}$$

where $P(\beta_i)$ is a polynomial; assumption iii makes (3.36) true at $\beta\chi = 0$.

$$\left| \int_{\mathcal{S}_g} \prod_{a=1}^l d\alpha_a \exp(-\sum_{a=1}^l \alpha_a m_a^2) \sum_{N' \in \Gamma} \prod_{\mathcal{S}' \in N} (-\tau^{-2l(\mathcal{S}) + c(\mathcal{S})}) Z(\alpha) \right| < 2^l$$

$$\times \sum_{a_j^i = 0}^{m(K_j^i)} \sum_{d_B = 0}^{m(B)} \int_0^\infty d\beta_i \, \beta_i^{p_i + 2l - 1} \exp(-\beta_i^2 m_{i_i}^2) P(\beta_i) \int_0^1 \prod_{i=0}^{l-1} (d\beta_i \, \beta_i^{p_i + 2i - 1}) \int_0^1 \prod_{\substack{i,j \\ m(H_j^i) \geqslant 0}} \left(d\chi_j^i \frac{(1 - \chi_j^i)^{m(H_j^i)}}{m(H_j^i)!} \right). \tag{3.37}$$

All the integrals in (3.37) exist because of Lemma II 5.

The proof of the theorem is then completed by noting that

$$\bigcup_g \mathcal{S}_g = \{\alpha : 0 \leqslant \alpha < \infty\}, \quad \text{and that} \quad \bigcup_{\Gamma \in W/R(g)} = W$$

(the sum over all equivalence classes of nests is the sum over all nests).

IV. APPLICATIONS TO FEYNMAN AMPLITUDES

A. Feynman amplitudes

The integrand of a Feynman amplitude, in a Euclidian space of dimension D, and expressed in the Schwinger

representation, as defined in Refs. 4, 7, and 10, is

$$Z(\alpha)=\left\{\pi\left(\frac{-1}{\sqrt{\alpha_a}}\frac{\partial}{\partial\delta_a}\right)\exp\left(+\sum_{a=1}^{l}\frac{\delta_a^2}{4}\right)\exp\left[-\sum_{i,j=1}^{n-1}\left(P_i+\sum_{a=1}^{l}\frac{\epsilon_{ia}\delta_a}{2\sqrt{\alpha_a}}\right)[d_G^{-1}(\alpha)]_{ij}\left(P_j+\sum_{b=1}^{l}\frac{\epsilon_{jb}\delta_b}{2\sqrt{\alpha_b}}\right)\right][P_G(\alpha)]^{-D/2}\right\}_{\delta_a=0}. \qquad (4.1)$$

In (4.1), the derivatives $\partial/\partial\delta$ take care of Feynmann amplitudes with spin or derivative couplings. The functions $P_G(\alpha)$ and $[d_G^{-1}(\alpha)]_{ij}$ as well as the matrix ϵ_{ia} are characteristic of the topology of the graph. In the exponential we sum over the vertices i and all lines a, as well as over the vectoriel indices of the Euclidian space.

It is important to note that $Z(\alpha)$ does have the "simultaneous Taylor series property" with respect to every nest N of subgraphs.[10] Then, by application of the theorem of absolute convergence, the renormalized amplitude is defined as

$$I_G(p,m)=\int_0^\infty\prod_{a=1}^{l}d\alpha_a\exp(-\sum_{a=1}^{l}\alpha_a m_a^2)R\{Z(\alpha)\}. \qquad (4.2)$$

It can be shown that the $\partial/\partial\delta$ derivatives can be taken through the R operator and outside the α integrals. The R operator is independent of the topology of the graph.

B. Oversubtractions (O.S.) rule

In Sec. III, the operator R is defined as

$$(1+\sum_N\prod_{S\in N}(\tau^{-2l(S)+c(S)})),$$

where the sum runs over all nests N of subgraphs. Then the theorem of absolute convergence holds if the over-subtractions coefficients $C(S)$ are non-negative and satisfy $C(S_1)\geq C(S_2)$ if $S_1\supset S_2$. If now, we impose to the non-negative coefficients $C(S)$ to satisfy the stronger condition

$$C(S_1\cup S_2)\geq C(S_1)+C(S_2)-C(S_1\cap S_2) \qquad (4.3)$$

for any subgraph S_1 and S_2, it is shown in Appendix B, that R can also be written in any order as $\prod_{S\subseteq G}(1-\tau^{-2l(S)+c(S)})$, where the product runs over the (2^l-1) subgraphs of G.

Definitions: Two subgraphs S_1 and S_2 are said to overlap if they have at least one line or one vertex in common and if $S_1\not\subset S_2$ or $S_2\not\subset S_1$.

A forest $\mathcal{F}$ is a set of non overlapping subgraphs.

A set of subdiagrams is said to be misjoint if they have no lines in common and if the number of loops in their union equals the sum of the number of loops in the individual subdiagrams.

An extended forest $\mathcal{E}$ is a set of subdiagrams such that any subset of mutually noninclusive elements is misjoint.

Now, if the rule (4.3) is satisfied for S_1 and S_2 disjoint, the operator R can also be expressed as

$$(1+\sum_{\mathcal{F}}\prod_{S\in\mathcal{F}}(\tau^{-2l(S)+c(S)})),$$

where we sum over all forests; if the rule (4.3) is satisfied for S_1 and S_2 misjoint, the operator R can be expressed as

$$(1+\sum_{\mathcal{E}}\prod_{S\in\mathcal{E}}(\tau^{-2l(S)+c(S)})),$$

where we sum over all extended forests. These definitions for R are now dependent of the topology of the graph.

Definition: A subgraph S is said to be one-vertex (one-line) reducible if there exists a vertex (line) such that upon its removal the number of connected parts of S increases. When R is expressed as

$$(1+\sum_X\prod_{S\in X}(-\tau^{-2l(S)+c(S)})),$$

the sum over X can be reduced to the sum over all (extended) forests of one-line irreducible subgraphs if the oversubtraction coefficients $C(S)$ satisfy

$$C(S)\leq C(S'), \qquad (4.4)$$

where S is a one-line reducible subgraphs and S' is its maximal one-line irreducible component. The sum over X can be further reduced to either a sum over all forests of connected, one line irreducible subgraphs, or a sum over all extended forests of connected, one-vertex irreducible subgraphs; such a reduction can be achieved if the oversubtraction coefficients $C(S)$ satisfy

$$C(\bigcup_{i=1}^{n}S_i)\leq\sum_{i=1}^{n}C(S_i), \qquad (4.5)$$

where any two subgraphs S_i have no common lines (for extended forests), or no common lines and vertices (for forests).

Definition: A subgraph S is said to be divergent (convergent) if

$$\omega(S)=L(S)D-2l(S)+d(S)+C(S) \qquad (4.6)$$

is positive or nul (negative). In (4.6), $L(S)$, $l(S)$, and $d(S)$ are successively the number of loops, lines and derivative couplings of S.

The sum over X is then further reduced either to a sum over all forests of connected, one-line irreducible, divergent subgraphs, or a sum over all extended forests of connected, one-vertex irreducible, divergent subgraphs.

Definition: A generalized vertex is a connected, one line irreducible subgraph S such that any other subgraph with the same vertices as those of S is contained in S.

If S_1 is a one-vertex irreducible, connected subgraph but is not a generalized vertex and if, given S_2 as the generalized vertex with the same vertices as S_1, the oversubtraction coefficients satisfy $C(S_2)\geq C(S_1)$, then the sum over X can be further reduced to a sum over all forests of diverging generalized vertices.

Those theorems are proved in Appendix B and are already valid in the special case of minimal subtractions ($C(S)=0$). The purpose of the O.S. rule is to preserve

this structure of the R operator when considering over-subtractions. The condition $C(S_1) \geq C(S_2)$ if $S_1 \supset S_2$ is compatible with (4.4) only if $C(S) = 0$, when $L(S) = 0$. Taking into account that (4.3) and (4.5) are compatible only if we impose the equal sign in (4.3) and (4.5) whenever S_1 and S_2 are disjoint (or misjoint), we can state the O.S. rule as follows:

$$C(S) \geq 0, \qquad (4.7a)$$

$$C(\phi) = C(S) = 0 \quad \text{if } L(S) = 0, \qquad (4.7b)$$

$$C(S) \geq C(S') \quad \text{for } S \supset S', \qquad (4.7c)$$

$$C(S_1 \cup S_2) = C(S_1) + C(S_2) \quad \text{for } S_1 \text{ and } S_2 \text{ disjoint} \qquad (4.7d)$$

The above rule is such that we preserve both, the absolute convergence of the oversubtracted Feynman amplitude, and the recurrence property of the R operation as stated by BPH[1,2,11] in terms of generalized vertices. Then, the R operator must be expressed as a sum over nests or a sum over forests. If in addition, we want to express the R operator as a product of $(1 - \tau)'_s$ (for practical computation), or as a sum over extended forests (useful for infrared problems[7]), we need to replace (4.7c) and (4.7d) by the stronger condition

$$C(S_1 \cup S_2) \geq C(S_1) + C(S_2) - C(S_1 \cap S_2), \qquad (4.7e)$$

with the equality sign required when S_1 and S_2 are misjoint.

We now give two examples of oversubtractions which were described recently in the literature. The oversubtracted normal products (isotropic or nonisotropic) are defined through a subtraction scheme based on forests.

Isotropic normal products[3]: An isotropic normal product is denoted by $N_\delta[P(\varphi)]$, where $P(\varphi)$ is a monomial of the fields φ and its derivatives and δ is an integer larger or equal to the degree d of the polynomial P (d is the number of factors of φ plus the number of derivatives ∂_μ). A Feynman amplitude derived from the time-ordered product of many normal products

$$\langle T[N_{\delta_1}[P_1(\varphi)](x_1) \cdots N_{\delta_n}[P_n(\varphi)](x_n)] \rangle^L \qquad (4.8)$$

is defined by the following assignment of oversubtractions

$$C(\phi) = C(S) = 0 \quad \text{if } L(S) = 0,$$

$$C(S) = \sum_{x_i \in S} (\delta_i - d_i), \qquad (4.9)$$

for S one-line irreducible and connected. Then the number of subtractions of the Feynman amplitude due to the subgraph S is given by the superficial degree of divergence of $S + C(S) + 1$ if this sum is > 0. The addition of $(\delta_i - d_i)$ subtractions coming from the vertex x_i is independent on which lines of the ith normal product are contained in S. Thus, it is called isotropic normal product.

The choice (4.9) for the coefficients $C(S)$ satisfies the O.S. rule [(4.7a)–(4.7d)] but fails to satisfy (4.7e) as it can be seen by taking S_1 and S_2 misjoint with x_i for common vertex.

Nonisotropic normal products[8]: Consider the ordered pair $\lfloor (f_1, f_2, \ldots, f_n), m \rfloor$, where $(f_1, \ldots, f_n)$ is a sequence of fields φ and their derivatives and m is a function that maps every subset of the set I_n of integers $1, 2, \ldots, n$, to a nonzero integer satisfying

$$m(\phi) = 0, \qquad (4.10a)$$

$$m(R) = 0, \qquad (4.10b)$$

if R is a subset of I_n containing only one integer,

$$m(R) \geq m(T) \quad \text{if } R \supseteq T. \qquad (4.10c)$$

Then, this ordered pair defines a non-isotropic normal product by the following two rules:

(i) The lines of this normal product in a Feynman diagram are $f_1, f_2, \ldots, f_n$ for a given vertex x_i.

(ii) The oversubtractions $C(S)$ of a one-line irreducible, connected subgraph S is defined by

$$C(S) = \sum_{x_i \in S} m_i[R_i(S)], \qquad (4.11)$$

where $R_i(S)$ is the set of the indices of those fields, in the normal product at the point x_i, that lie in the subgraph S. Let us mention that such anisotropic normal products are used to prove the equivalence theorem[12] under a field transformation $\varphi \to \varphi + F(\varphi)$.

The oversubtraction rules [(4.10a)–(4.10c)] satisfy the O.S. rules [(4.7a)–(4.7d)]. To satisfy (4.7e) it is necessary to change (4.10c) into the stronger rule

$$m(R_1 \cup R_2) \geq m(R_1) + m(R_2) - m(R_1 \cap R_2), \qquad (4.10d)$$

with the equal sign required whenever $R_1 \cap R_2 = \phi$.

C. Some other examples

Soft mass subtractions.[9] The function

$$Z(\alpha) = \exp\left[-\left(\sum_{a=1}^{l} \alpha_a \mu_a^2 + \sum_{i,j=1}^{n-1} p_i [d_G^{-1}(\alpha)]_{ij} p_j \right) \right] [P_G(\alpha)]^{-D/2} \qquad (4.12)$$

also has the "simultaneous Taylor series property" with respect to every nest N of subgraphs. The renormalized scalar amplitude with soft μ mass

$$I_G(p, m, \mu) = \int_0^\infty \prod_{a=0}^{l} d\alpha_a \exp\left(-\sum_{a=1}^{l} \alpha_a m_a^2 \right)$$

$$\times R\left\{ \exp\left(-\sum_{a=1}^{l} \alpha_a \mu_a^2 \right) \right.$$

$$\times \exp\left(-\sum_{i,j=1}^{n-1} p_i [d_G^{-1}(\alpha)]_{ij} p_j \right) [P_G(\alpha)]^{-D/2} \qquad (4.13)$$

is absolutely convergent in Euclidean space. This result can be extended to the case of derivative couplings.

Mellin transform of a Feynman amplitude.[6,7] The Mellin transform of a renormalized scalar Feynman amplitude in a Euclidian space of dimension D is

$$M_G(p, m, x) = \tfrac{1}{2} \Gamma\left(-\frac{x}{2} \right) \int_0^\infty \prod_{a=1}^{l} d\alpha_a \exp\left(-\sum_{a=1}^{l} \alpha_a m_a^2 \right)$$

$$\times R\left\{ \left(\sum_{i,j=1}^{n-1} p_i [d_G^{-1}(\alpha)]_{ij} p_j \right)^{x/2} [P_G(\alpha)]^{-D/2} \right. \qquad (4.14)$$

This integral is absolutely convergent for any complex x. This result can be extended to the case of derivative couplings.

V. CONNECTION WITH BPH

In this section, we intend to show that the R operator introduced in Sec. III and used in Sec. IV A, fulfill the recursive solution to the problem of renormalization as given by Bogolubov, Parasiuk, and Hepp.[1,2,11] Thus, such an R operator can be used to construct in perturbation, a field theory which is Poincaré invariant, unitary, and causal. For simplicity, we remain in Euclidian space since the algorithm of the R operation is similar in Euclidian and in Minkowsky space.

Let us define the Feynman amplitude with cutoff $r > 0$,

$$\bar{R}_G^r(p,m) = \int_r^\infty \prod_{a=1}^l d\alpha_a \exp(-\sum_{a=1}^l \alpha_a m_a^2)$$

$$\times (1 + \sum_N \prod_{S \in N} (-\tau^{-2l(S)+c(S)})) Z_G(\alpha), \tag{5.1}$$

where $Z_G(\alpha)$ is defined in (4.1) and is characteristic of a Feynman amplitide with derivative couplings and spinors. Then, in Sec. III, we have seen that

$$\lim_{r \to 0} \tilde{R}_G^r(p,m) = I_R(p,m). \tag{5.2}$$

Now, in Appendix B 5, we show that the R operator can also be written as a sum over forests of diverging generalized vertices if the oversubtraction coefficients $C(S)$ satisfy the O.S. rule [(4.7a)–(4.7d)].

Let $\overline{U}^{(G)}$ be the set of all those forests of diverging generalized vertices of G that do not contain G. Then, because of the cutoff $r > 0$, we find

$$\tilde{R}_G^r(p,m) = \bar{\bar{R}}_G^r(p,m) + \tilde{\chi}_G^r(p,m), \tag{5.3}$$

where

$$\bar{\bar{R}}_G^r(p,m) = \int_r^\infty \prod_{a=1}^l d\alpha_a \exp(-\sum_{a=1}^l \alpha_a m_a^2)$$

$$\times (1 + \sum_{J \in \overline{U}^{(G)}} \prod_{S \in J} (-\tau^{-2l(S)+c(S)})) Z_G(\alpha), \tag{5.4}$$

and

$$\tilde{\chi}_G^r(p,m) = \int_r^\infty \prod_{a=1}^l d\alpha_a \exp(-\sum_{a=1}^l \alpha_a m_a^2)$$

$$\times (1 + \sum_{J \in \overline{U}^{(G)}} \prod_{S \in J} (-\tau^{-2l(S)+c(S)}))$$

$$\times (-\tau_G^{-2l(G)+c(G)}) Z_G(\alpha). \tag{5.5}$$

From (3.1), (3.2), and (4.1), we see that the τ operator acting upon the α's of the entire graph G is equivalent to a Taylor T operator acting upon the external legs of the entire graph G,

$$\tau_G^{-2l(G)+c(G)} Z_G(\alpha) = T_{p_i}^{\omega(G)+c(G)} Z_G(\alpha), \tag{5.6}$$

where

$$\omega(G) = L(G)D - 2l(G) + d(G), \tag{5.7}$$

and where $L(G)$, $l(G)$, and $d(G)$ are, respectively, the

number of loops, lines, and derivative coupling of G. Then

$$\tilde{\chi}_G^r(p,m) = -T_{p_i}^{\omega(G)+c(G)} \bar{\bar{R}}_G^r(p,m). \tag{5.8}$$

By Fourier transform of (5.3) and (5.8), we obtain

$$R_G^r(x,m) = \bar{R}_G^r(x,m) + \chi_G^r(x,m), \tag{5.9}$$

with

$$\chi_G^r(x,m) = -M \bar{R}_G^r(x,m), \tag{5.10}$$

and where the operator M is defined by (5.8) and the Fourier transformation. If x in (5.10) means n vertices $(x_1, \ldots, x_n)$, $\chi_G^r(x,m)$ is a quasilocal object of the form $Z(\partial/\partial x_k)\delta(x_1 - x_n) \cdots \delta(x_{n-1} - x_n)$.

In the expression for $\bar{R}_G^r(p,m)$, we may sum over those forests having common maximal generalized vertices, and then sum over the set of these maximal elements

$$R_G^r(p,m) = \sum_{\{R_1,\ldots,R_t\}} \int_r^\infty \prod_{a=1}^l d\alpha_a \exp(-\sum_{a=1}^l \alpha_a m_a^2)$$

$$\times \prod_{i=1}^t [(1 + \sum_{J \in \overline{U}(R_i)} \prod_{S \in J} (-\tau^{-2l(S)+c(S)}))$$

$$\times (-\tau_{R_i}^{-2l(R_i)+c(R_i)})] Z_G(\alpha), \tag{5.11}$$

where $\{R_1, \ldots, R_t\}$ is a set of disjoint divergent generalized vertices.

The application of each $\tau_{R_i}^{-2l(R_i)+c(R_i)}$ operator upon the function $Z(\alpha)$ gives an expression of the form[7,10]

$$-\tau_{R_i}^{-2l(R_i)+c(R_i)} Z_G(\alpha)$$

$$= -\sum_{q=0}^{\omega(R_i)+c(R_i)} \frac{1}{q!} \left(\frac{\partial}{\partial k_{i_1}} \cdots \frac{\partial}{\partial k_{i_q}} Z_{R_i}(\alpha) \right)_{k_i=0}$$

$$Z_{(G/R_i)}(k_{i_1},\ldots,k_{i_q}), \tag{5.12}$$

where $[G/R_i]$ denotes the reduced graph obtained from G by shrinking R_i into a point; the momenta k_i's are external momentum to the subgraph R_i and are a sum of external and internal momentum to the graph G. In the α representation, a derivative coupling on an internal line a is denoted by $(-1/\sqrt{\alpha_a})\partial/\partial\delta_a$ [see (4.1)]. Then,

$$k_j = p_j + \sum_{a \in [G/R_i]} \frac{\epsilon_{ja}}{\sqrt{\alpha_a}} \frac{\partial}{\partial \delta_a}, \tag{5.13}$$

where ϵ_{ja} is the incidence matrix of the graph G. Finally,

$$Z_{(G/R_i)(k_{i_1},\ldots,k_{i_q})}(\alpha)$$

$$= \left\{ \prod_{j=1}^9 \left(p_{ij} + \sum_{a \in [G/R_i]} \frac{\epsilon_{ja}}{\sqrt{\alpha_a}} \frac{\partial}{\partial \delta_a} \right) Z_{(G/R_i)} \left(\alpha, \frac{\partial}{\partial \delta}, \delta \right) \right\}_{\delta_a=0}, \tag{5.14}$$

with $Z_G(\alpha, \partial/\partial\delta, \delta)$ given by the inside of the curly bracket $\{\}$ in (4.1). In (5.12), summation over all possible external momenta k_i's and over vector indices is understood.

From the definition (5.4), we get

$$\int_r^\infty \prod_{a\in R_i} d\alpha_a \exp(-\sum_{a\in R_i}\alpha_a m_a^2)$$

$$\times(1+\sum_{j\in \bar U(R_i)} S\bigsqcup_{S\in J}(-\tau^{-2I(S)+C(S)}_S))$$

$$\times(-\tau^{-2I(R_i)+C(R_i)})Z_G(\alpha)$$

$$=-\sum_{q=0}^{\omega(R_i)+C(R_i)}\frac{1}{q!}\left(\frac{\partial}{\partial k_{i_1}}\cdots\frac{\partial}{\partial k_{i_q}}\widetilde{R}_{R_i}^r(k,m)\right)_{k_i=0}$$

$$\times Z_{[G/R_i]}(k_{i_1},\ldots,k_{i_q})}(\alpha). \tag{5.15}$$

Using (5.8), we can write $\widetilde{R}_G^r(p,m)$ as

$$R_G^r(p,m)=\sum_{\{R_1,\ldots,R_t\}}\int_r^\infty \prod_{a\in[G/UR_i]} d\alpha_a$$

$$\times\exp(-\sum_{a\in[G/UR_i]}\alpha_a m_a^2)$$

$$\times\left\{\prod_{i=1}^t \widetilde{\chi}_{R_i}^r\left(p_j+\sum_{a\in[G/R_i]}\frac{\epsilon_{ja}}{\sqrt{\alpha_a}}\frac{\partial}{\partial\delta_a},m\right)\right.$$

$$\left.\times Z_{[G/UR_i]}(\alpha,\frac{\partial}{\partial\delta},\delta)\right\}_{\delta_a=0}, \tag{5.16}$$

where UR_i means the union of all the subgraphs R_i.

To calculate $\bar R_G^r(x,m)$ in position space, it is convenient to remind the reader what the curly bracket $\{\ \}_{\delta_a}$ $\{\ \}_{\delta_a=0}$ in (5.16) is in momentum space,[10]

$$\{\ \}_{\delta_a=0}\sim\int_{-\infty}^{+\infty}\prod_{a\in[G/UR_i]} d^D k_a$$

$$\times\prod_{i=1}^t \widetilde{\chi}_{R_i}^r(p_j-\sum_{a\in[G/R_i]}\epsilon_{ja}k_a,m)\prod_{a\in[G/UR_i]}$$

$$\times (k_a,s\exp(-\alpha_a k_a^2))$$

$$\times\prod_{n\in[G/UR_i]}\delta^{(D)}(p_n-\sum_{a\in[G/UR_i]}\epsilon_{na}k_a), \tag{5.17}$$

where the k_a's take care of possible derivative couplings or spinor on the line a and the $\delta^{(D)}$ distributions describe energy momentum conservation at each vertex n. The right-hand side of (5.17) is a convolution and its Fourier transform is the product of the t functions $\chi(x,m)$ times the Feynman amplitude for the reduced graph $[G/UR_i]$, where at each contracted point R_i, the external momentum is $\sum_{j\in R_i}p_j$. Consequently,

$$\bar R_G^r(x,m)=\sum_{\{R_1,\ldots,R_t\}}\prod_{i=1}^t \chi_{R_i}^r(x,m)$$

$$\times\prod_{a\in[G/UR_i]}\Delta_F^r(\sum_{n\in[G/UR_i]}\epsilon_{na}x_n), \tag{5.18}$$

where $\{R_1,\ldots,R_t\}$ is a set of disjoint divergent generalized vertices and Δ_F^r are Feynman-like propagators with ultraviolet cutoff r. Equations (5.9), (5.10), and

(5.18) define the recursive solution to the R-operation as given by Bogolubov, Parasiuk, and Hepp.[1,2,11]

VI. CONCLUSION

Most problems on renormalization were solved in the past by using recurrence arguments. In 1973, Zimmermann[3] introduced a renormalized Feynman amplitude written in compact form in the momentum representation. This new result was the starting point for the normal product algorithm.[3] In the same spirit, this paper defines a renormalized Feynman amplitude written in compact form in Schwinger parametric representation. Let us mention some features of this representation: the subtraction R operator is independent of the topology of the graph; the generalized Taylor operator τ relative to subgraphs which form a forest commute; we avoid completely the problem which occur in momentum representation of defining a permissible set of internal momenta. The R operator turns out to be easy to manipulate for practical computations like Zimmermann's identity,[10] calculation of the coefficients of all logarithms of the leading power in the asymptotic behavior of a renormalized amplitude,[7] infrared and ultraviolet convergent amplitude, etc.

Finally, let us remind the basic property of Feynman amplitudes expressed in the α representation which makes such a R operator to exist: the α-Feynman integrand has a simultaneous Laurent series in the dilatation variables corresponding to subgraphs which form a forest.

ACKNOWLEDGMENTS

APPENDIX A: GENERALIZED TAYLOR OPERATORS

The generalized Taylor operators have been defined and extensively employed in Ref. 4. Here we want to further generalize the definition over a class of functions $f(x)$ which behaves like x^ν at $x=0$ (where ν is not necessarily an integer).

Definition: Given a function $f(x)$ such that $x^{-\nu}f(x)$ is C^∞ in $[0,a>0,]$ we define the generalized Taylor operator τ^n on $f(x)$ as

$$\tau^n f(x)=x^{-\lambda-\epsilon}T^{n+\lambda}\{x^{\lambda+\epsilon}f(x)\}, \tag{A1}$$

where $\lambda\geq -E'(\nu)$ is an integer, $E'(\nu)$ is the smallest integer $\geq\mathrm{Re}\,\nu$, and $\epsilon=E'(\nu)-\nu$. In (A1) n is an integer and T is the usual Taylor operator.

The above definition is λ independent. The purpose of introducing ϵ is to remove the cut of $f(x)$ at $x=0$. Let us mention the following properties:

$$\tau^n f(x)\sim x^q \quad \text{at } x\sim 0 \quad \text{with } \mathrm{Re}\,q\leq n, \tag{A2}$$

$$(1-\tau^n)f(x)\sim x^q \quad \text{at } x\sim 0 \quad \text{with } \mathrm{Re}\,q>n, \tag{A3}$$

$$\tau^n f(x)=0 \quad \text{if } n-E'(\nu)<0. \tag{A4}$$

By using the formula for the remainder of the Taylor expansion, we have

$$(1 - \tau^n)f(x) = \int_0^1 d\xi \, \frac{(1 - \xi)^{n+\lambda}}{(n+\lambda)!} \, \frac{\partial^{n+\lambda+1}}{\partial \xi^{n+\lambda+1}} \{\xi^{\lambda+\epsilon} f(x\xi)\}. \quad (A5)$$

The ϵ in (A5) is essential to ensure the existence of the integral. In this equation, $\lambda \geq \sup(-E'(\nu), -n)$.

The generalization of the definition of τ to functions of several variables is straightforward but in general the the τ operators do not commute. We observe the properties

$$\tau_x^n \tau_y^n f(x,y)$$

$$\sim \begin{cases} x^q & \text{for } x \sim 0, \ y \neq 0, \ \text{and } \mathrm{Re}q \leq n_x, \\ y^q & \text{for } y \sim 0, \ x \neq 0, \ \text{and } \mathrm{Re}q \leq n_y, \end{cases} \quad (A6)$$

$$(1 - \tau_x^n)(1 - \tau_y^n)f(x,y) \sim x^q \quad \text{for } x \sim 0, \ y \neq 0, \ \text{and } \mathrm{Re}q > n_x, \quad (A7)$$

but nothing can be said on the behavior at $y \sim 0$, $x \neq 0$.

$$\cdots (1 - \tau_{x_i}^{n_i}) \cdots \tau_{x_i}^{n_i'} \cdots f = 0 \quad \text{if } n'_i \leq n_i. \quad (A8)$$

Corollary

$$\cdots (1 - \tau_{x_i}^{n_i}) \cdots f = \cdots (1 - \tau_{x_i}^{n_i}) \cdots (1 - \tau_{x_i}^{n_i'}) \cdots f \quad \text{if } n'_i \leq n_i, \quad (A9)$$

$$\cdots \tau_{x_i}^{n_i'} \cdots f = \cdots \tau_{x_i}^{n_i} \cdots \tau_{x_i}^{n_i'} \cdots f \quad \text{if } n'_i \leq n_i, \quad (A10)$$

where $\cdots$ means a sequence of τ operators. The integral representation for the remainder of the Taylor expansion is not always generalizable to functions of several variables. Indeed in $(1 - \tau_y)(1 - \tau_x)f$, ϵ_y is generally different for each term in $(1 - \tau_x)f$.

Finally, let us define the τ operators relative to a family of variables. Given a function of several variables $f(\{x\}, \{y\})$, where $\{x\}$ and $\{y\}$ are families of variables, we define

$$\tau_{\{x\}}^n f(\{x\}, \{y\}) = [\tau_\rho^n f(\{\rho x\}, \{y\})]_{\rho=1}. \quad (A11)$$

APPENDIX B: DIFFERENT FORMS OF THE R OPERATOR

In Sec. III, for the proof of absolute convergence, it was convenient to use the R operator as a sum over all nests N of subgraphs. Here we prefer to define the R operator under its original form (see Ref. 4)

$$RZ(\alpha) = \prod_{S \subseteq G} (1 - \tau_S^{-2l(S)+C(S)})Z(\alpha), \quad (B1)$$

where the product runs over $2^l - 1$ subgraphs of G. This form is independent of the topology of the graph and $RZ(\alpha)$ turns out to be independent of the order of application of the τ operators.

1. The nested forest formula

Theorem: If $Z(\alpha)$ has the "simultaneous Taylor series property" in regards to every nest,

$$\prod_{S \subseteq G} (1 - \tau_S^{-2l(S)+C(S)})Z(\alpha)$$

$$= (1 + \sum_N \prod_{S \in N} (-\tau_S^{-2l(S)+C(S)}))Z(\alpha), \quad (B2)$$

provided that the oversubtraction coefficients $C(S)$ satisfy for any two subgraphs S_1 and S_2,

$$C(S_1 \cup S_2) \geq C(S_1) + C(S_2) - C(S_1 \cap S_2). \quad (B3)$$

Proof: The proof is by recurrence. We consider

$$\prod_{l=1}^{2^l-1} (1 - \tau_{S_l}^{-2l(S_l)+C(S_l)})Z(\alpha) = \prod_{l=m}^{2^l-1} (1 - \tau_{S_l}^{-2l(S_l)+C(S_l)})$$

$$\times (1 + \sum_{N \in \mathcal{E}_{m-1}} \prod_{S \in N} (-\tau_S^{-2l(S)+C(S)}))Z(\alpha), \quad (B4)$$

where $\mathcal{E}_{m-1}$ is the set of all forests of nested elements built from the subdiagrams in $W_{m-1} = \{S_1, \ldots, S_{m-1}\}$. For $m = 2$, it is trivially true, while for $m = 2^l$, it reduces to (B2). If we assume it to be valid for $m = n - 1$, then it is valid for $m = n$ provided that

$$\prod_{l=n+1}^{2^l-1} (1 - \tau_{S_l}^{-2l(S_l)+C(S_l)})(-\tau_{S_n}^{-2l(S_n)+C(S_n)})$$

$$\times \sum_{N \in \mathcal{E}'_{n-1}} \prod_{S \in N} (-\tau_S^{-2l(S)+C(S)})Z(\alpha) = 0, \quad (B5)$$

where $\mathcal{E}'_{n-1}$ is the set of all forests of nested elements built from the subdiagrams $S_1, \ldots, S_{n-1}$ with at least one element either disjoint or overlapping with S_n. Hence, the nested forest formula (B2) is proved by establishing (B5).

We use the Ω construction of S_n-maximal nest $\mathcal{G}$. To each nest $N \in \mathcal{E}'_{n-1}$ corresponds a S_n-maximal nest $\mathcal{G}$. $\mathcal{G}$ can be decomposed into B, K, and H and every nest N with S_n-maximal nest $\mathcal{G}$ is of the form $B \cup K \cup H'$, with $H' \subseteq H$. Now, some elements of H do not belong to $\mathcal{E}'_{n-1}$; when we group the nests of $\mathcal{E}'_{n-1}$ into equivalent classes, we obtain from the left-hand side of (B5), for each equivalent class

$$\prod_{l=n+1}^{2^l-1} (1 - \tau_{S_l}^{-2l(S_l)+C(S_l)})(-\tau_{S_n}^{-2l(S_n)+C(S_n)})$$

$$\times \prod_{S \in B \cup K} (-\tau_S^{-2l(S)+C(S)}) \prod_{S \in H \cap W_{n-1}} (1 - \tau_S^{-2l(S)+C(S)})Z(\alpha). \quad (B6)$$

Then, using the property (A9) of the τ operators, (B6) becomes

$$\prod_{l=n+1}^{2^l-1} (1 - \tau_{S_l}^{-2l(S_l)+C(S_l)})(-\tau_{S_n}^{-2l(S_n)+C(S_n)})$$

$$\times \prod_{S \in B \cup K} (-\tau_S^{-2l(S)+C(S)}) \prod_{S \in H} (1 - \tau_S^{-2l(S)+C(S)})Z(\alpha). \quad (B7)$$

Now, the proof is similar to the proof of absolute convergence; we dilate the α's in B_s by ξ_s^2, the α's in K_j by σ_j^2, the α's in H_j by χ_j^2 and the α's in S_n by β^2. Then, $Z(\alpha)$ becomes $Z(\alpha, \xi_s, \sigma_j/\beta, \beta\chi_j)$. After application of the τ operators in $B \cup K$, and after using the remainder formula for the elements of H, we obtain a sum of terms each of them containing in factor the term β^p, with

$$p > \sum_{j=1}^{r-1} [2l(K_j) - 2l(H_j) + C(H_j) - C(K_j)]. \quad (B8)$$

Now, if the oversubtraction coefficients satisfy the re-

lation (B3) then, by the same derivation as (3.24), we get

$$\sum_{j=1}^{r-1} [C(H_j) - C(K_j)] \geq C(S_n),$$
(B9)

so that

$$p > -2l(S_n) + C(S_n).$$
(B10)

Then, the application of $\tau_{S_n}^{-2l(S_n)+c(S_n)}$ in (B7) gives zero by (A4), and this proves (B5) and consequently (B2).

This theorem shows that if the condition (B3) over the oversubtraction coefficients is satisfied, the R operator used in Sec. III, namely,

$$RZ(\alpha) = (1 + \sum_N \prod_{S \in N} (-\tau_S^{-2l(S)+c(S)}))Z(\alpha),$$
(B11)

is the same as the R operator defined in (B1). This property is true without any reference to Feynman graphs.

2. Forest formula

From now on, the function $Z(\alpha)$ is related to a Feynman graph. We shall generalize the nested forest formula for those functions $Z(\alpha)$ which have Taylor series in the dilatation variables corresponding to the diagrams of some forests in addition to the forests of nested elements, after all common factors have been removed. (Definition of a forest given in Sec. IV B).

Let us note that any forest which is not a nest has some disjoint elements. Given a forest, a set of disjoint elements of this forest is said to be maximal if any element of the forest that does not contain all of them, is contained in one of them. If no such set exists, then the forest is a forest of nested elements. We group all forests which are not a nest into pairs of the form

$$\{S_1, \ldots, S_n, \text{rest}\} \text{ and } \{S_1 \cup \cdots \cup S_n, S_1, \ldots, S_n, \text{rest}\},$$
(B12)

where $\{S_1, \ldots, S_n\}$ is the maximal disjoint set of elements and where of course $S_1 \cup \cdots \cup S_n$ does not belong to the rest of the forest. Either $Z(\alpha)$ has a Taylor series property with respect to both forests of the pair or with respect to none. If $Z(\alpha)$ has a Taylor series property, the τ's relative to the elements of these forests commute and we can form the sum

$$\circ \circ \circ (1 - \tau_{\cup S_i}^{-2l(\cup S_i)+c(\cup S_i)})(-\tau_{S_1}^{-2l(S_1)+c(S_1)}) \cdots$$
$$\times (-\tau_{S_n}^{-2l(S_n)+c(S_n)})Z(\alpha),$$
(B13)

which vanishes by virtue of (A2) and (A3) if $C(\cup S_i) \geq \sum_{i=1}^{n} C(S_i)$. Summing over all such pairs of forests and adding the result to the nested forest formula (B11), we obtain

$$\prod_{i=1}^{2^l-1} (1 - \tau_{S_i}^{2l(S_i)+c(S_i)})Z(\alpha) = (1 + \sum_J \prod_{S \in J} (\tau_S^{-2l(S)+c(S)}))Z(\alpha),$$
(B14)

where the sum runs over all forests which have the Taylor series property. Then, the left-hand side of (B14) is independent of the ordering.

The same demonstration applies to establish the

extended forest formula. In the above proof we just have to change the words "forest J" into the words "extended forest $\mathcal{E}$," "disjoint" into "misjoint" (definitions given in Sec. IV B).

3. (Extended) forest formula of (one-vertex) one-line irreducible connected subgraphs

From now on, we consider the function $Z(\alpha)$ given in (4.1). In this subsection, we want to reduce the sum in the right-hand side of (B14) into a sum over a subset of (extended) forests, each (extended) forest containing only connected, (one-vertex) one-line irreducible subgraphs as defined in Sec. IV B.

Theorem: The sum of all (extended) forests, each containing at least a one-line reducible subgraph, vanishes on $Z(\alpha)$ if the oversubtraction coefficients satisfy

$$C(S) \leq C(S'),$$
(B15)

where S is a one-line reducible subgraph and S' is its maximal one-line irreducible component.

Proof: Any such (extended) forest has at least one minimal one-line reducible subgraph S. Let S' be its maximal one-line irreducible component. We gather the following terms:

$$\cdots (1 - \tau_{S'}^{-2l(S')+c(S')})(-\tau_S^{-2l(S)+c(S)})Z(\alpha).$$
(B16)

After dilatation of the $\sqrt{\alpha}$'s belonging to S by ρ and to S' by μ, the function $Z(\alpha, \rho, \mu)$ can be written

$$Z(\alpha, \rho, \mu) = (\rho\mu)^{-k}Z'(\alpha, \rho\mu, \rho),$$
(B17)

where $Z'(\alpha, \rho\mu, \rho)$ has a Taylor expansion in $\rho\mu$ and ρ (this property is proved in Ref. 10).

Then, (B16) becomes

$$\cdots (1 - \tau_\mu^{-2l(S')+c(S'+k)})(-\tau_\rho^{-2l(S)+c(S)+k})$$
$$\sum_{k_1, k_2 = 0}^{\infty} Z_{k_1 k_2}(\rho\mu)^{k_1}\rho^{k_2}.$$
(B18)

Using (A2), (A3), and (B15), we see that the expansion (B18) vanishes.

Theorem: The sum of all (extended) forests, each containing at least (a one-vertex reducible or/and) a disconnected subgraph vanishes on $Z(\alpha)$ if the oversubtraction coefficients $C(S)$ satisfy

$$C(S) \leq \sum_{i=1}^{n} C(S_i),$$
(B19)

where the subgraphs S_i are the n (one-vertex irreducible) connected components of the subgraph S.

Proof: Any such (extended) forest has at least one minimal (one-vertex reducible or/and) disconnected subgraph S. Let us partition S into n (one-vertex irreducible) connected subgraphs $S_1, \ldots, S_n$, such that $S = \cup_{i=1}^{n} S_i$ and let us sum over all (extended) forests containing S but which differ only by the number of subgraphs S_i (i from 1 to n); we get

$$\cdots (1 - \tau_{S_1}^{-2l(S_1)+c(S_1)}) \cdots (1 - \tau_{S_n}^{-2l(S_n)+c(S_n)})$$
$$\times (-\tau_S^{-2l(S)+c(S)})Z(\alpha).$$
(B20)

1556   J. Math. Phys., Vol. 17, No. 8, August 1976

M.C. Bergere and Y.-M.P. Lam   1556

Since $Z(\alpha)$ has the simultaneous Taylor property in regards to the subgraphs $S_1, \ldots, S_n$ and S with common powers $\mu_1, \ldots, \mu_n$ and $\sum_{i=1}^{n} \mu_i$, the above expression vanishes by virtue of (A2), (A3), and (B19).

4. (Extended) forest formula of divergent subgraphs

Theorem: If S is a convergent subgraph (definition given in section IV B)),

$$\tau_S^{-2i(S)+c(S)} Z(\alpha) = 0. \tag{B21}$$

The proof is a direct application of (A4).

5. Forest formula of generalized vertices (definition given in Sec. IV B)

Theorem: The sum of all forests, each containing a subgraph which is not a generalized vertex, vanishes on $Z(\alpha)$ if the dimension of space $D \geq 2$ and if the over-subtraction coefficients $C(S)$ satisfy

$$C(S_2) \geq C(S_1) \text{ for } S_2 \supset S_1. \tag{B22}$$

Proof: Any such forest has at least one maximal subgraph S_1 which is not a generalized vertex. Let us add to S_1 n lines to obtain the generalized vertex S_2 such that S_2 and S_1 have the same vertices. S_2 has n lines and n loops more than S_1. We gather the following terms:

$$\cdots (1 - \tau_{S_2}^{-2i(S_2)+c(S_2)})(- \tau_{S_1}^{-2i(S_1)+c(S_1)}) Z(\alpha). \tag{B23}$$

After dilatation of the $\sqrt{\alpha}$'s belonging to S_1 by ρ and to S_2 by μ, the function $Z(\alpha, \rho, \mu)$ can be written

$$Z(\alpha, \rho, \mu) = \rho^{-L(S_1)D} \mu^{-L(S_1)D+nD} Z'(\alpha, \rho\mu, \rho), \tag{B24}$$

where $Z'(\alpha, \rho\mu, \rho)$ has a Taylor expansion in $\rho\mu$ and ρ (this property is proved in Ref. 10).

Then (B23) becomes

$$\cdots (1 - T_\mu^{-2i(S_2)+[L(S_1)+n]D+c(S_2)})(- T_\rho^{-2i(S_1)+L(S_1)D+c(S_1)})$$

$$\sum_{k_1, k_2=0}^{\infty} Z_{k_1 k_2} (\rho\mu)^{k_1} \rho^{k_2}. \tag{B25}$$

By (A2), (A3), and (B22), the above expression vanishes for $D \geq 2$.

Similar results as those of subsections 3, 4, and 5 of the appendix can be also obtained from the complete product of $(1 - \tau)$'s:

$$\prod_{S \subseteq G} (1 - \tau_S^{-2i(S)+c(S)}) Z(\alpha) = \prod_{S' \subseteq G} (1 - \tau_{S'}^{-2i(S')+c(S')}) Z(\alpha), \tag{B26}$$

where the product on the right hand side runs only over divergent generalized vertices. Such a reduced form turns out to be useful for practical computation of the renormalized integrand.

[1]N.N. Bogolubov and O. Parasiuk, Acta Math. **97**, 227 (1957).

[2]K. Hepp, Commun. Math. Phys. **2**, 301 (1966).

[3]W. Zimmermann, Ann. Phys. (N.Y.) **77**, 536 (1973).

[4]M.C. Bergère and J.B. Zuber, Commun. Math. Phys. **35**, 113 (1974).

[5]T. Appelquist, Ann. Phys. (N.Y.) **54**, 27 (1969); S.A. Anikin, M.K. Polivanov, and O.I. Zavialov, Dubna preprint E2-7433 (1973) (unpublished).

[6]M.C. Bergère and Y.M.P. Lam, Commun. Math. Phys. **39**, 1 (1974).

[7]M.C. Bergère and Y.M.P. Lam, "Asymptotic Expansion of Feynman Amplitudes, Part II: The divergent case," Freie Universitat Berlin, HEP May 74/9 (unpublished).

[8]Y.M.P. Lam, Phys. Rev. D **6**, 2145 (1972); **6**, 2161 (1972).

[9]M. Gomes, J.H. Lowenstein, and W.W. Zimmermann, Commun. Math. Phys. **39**, 1 (1974).

[10]M.C. Bergère, Y.M.P. Lam, Lectures on Renormalization Techniques in the α-Parametric Representation, Course given in Saclay (January 1975) unpublished.

[11]N.N. Bogolubov and D.W. Shirkov, *Introduction to the Theory of Quantized Fields* (Interscience, New York, 1959).

[12]Y.M.P. Lam, Phys. Rev. D **7**, 2943 (1973); M.C. Bergère and Y.M.P. Lam, Equivalence theorem and Fadeev-Popov Ghosts, Saclay preprint, June 1975, DPh. T/75/GS (unpublished).

On the structure of the multiplicity-free Wigner coefficients of U(n)

M. K. F. Wong

Fairfield University, Fairfield, Connecticut 06430
(Received 2 April 1976)

It is shown that the isoscalar factor (or reduced Wigner coefficient)

$$\begin{pmatrix} m_{1n} \cdots m_{n-1\,n} \; 0 & \Big\| p, \overset{\circ}{0} \Big\| & m'_{1n} \cdots m'_{n-1\,n} \; 0 \\ m_{1\,n-1} \cdots m_{n-1\,n-1} & \Big\| q, \overset{\circ}{0} \Big\| & m'_{1\,n-1} \cdots m'_{n-1\,n-1} \end{pmatrix}$$

in U(n) is essentially a doubly stretched 9-j symbol in U($n-1$). The connection between the isoscalar factor

$$\begin{pmatrix} m_{1n} \cdots m_{nn} & \Big\| p, \overset{\circ}{0} \Big\| & m'_{1n} \cdots m'_{nn} \\ m_{1\,n-1} \cdots m_{n-1\,n-1} & \Big\| q, \overset{\circ}{0} \Big\| & m'_{1\,n-1} \cdots m'_{n-1\,n-1} \end{pmatrix}$$

and the 9-j symbol of U(n) and U($n-1$) is also noted. This result immediately implies that the Weyl coefficients of U(n) are basically 6-j symbols of U($n-1$), a result first noted by Holman. The finite transformation matrix $D\,{}^{[m]_n}_{(m')_{n-1}(m)_{n-1}}$ either in terms of generalized Euler angles or double bosons can thus be written down in a simple way. The stretched 6-j symbols of U(n) are obtained in a simple form, involving no summations. The generalized beta functions of Gel'fand and Graev for U(n) are found to be connected with the stretched 6-j symbols of U($n-1$) and an isoscalar factor of U($n-1$). The 144 Regge symmetries of the 6-j symbol of U(2) can be interpreted as the symmetries of the Weyl coefficients of the double boson state of U(3) * U(3). In the Appendix we give the phase relations between the Wigner coefficients and 3-j symbols of U(n), a result which is by no means trivial, and is of some practical importance.

INTRODUCTION

Although the explicit expression of the multiplicity-free Wigner coefficients of U(n) has been known for some time,[1-8] it is still useful to study its structure so as to obtain more information from it, and also to recognize its symmetry properties. In this direction we would like to offer the following suggestion: The isoscalar factor

$$\begin{pmatrix} m_{1n} \cdots m_{n-1\,n} \; 0 & \Big\| p, \overset{\circ}{0} \Big\| & m'_{1n} \cdots m'_{n-1\,n} \; 0 \\ m_{1\,n-1} \cdots m_{n-1\,n-1} & \Big\| q, \overset{\circ}{0} \Big\| & m'_{1\,n-1} \circ \circ m'_{n-1\,n-1} \end{pmatrix},$$

where

$$p = \sum_{i=1}^{n-1} m_{in} - \sum_{i=1}^{n-1} m'_{in}, \quad q = \sum_{i=1}^{n-1} m_{i\,n-1} - \sum_{i=1}^{n-1} m'_{i\,n-1}$$

is connected with the 9-j symbols of U($n-1$), and the isoscalar factor

$$\begin{pmatrix} m_{1n} \cdots m_{nn} & \Big\| p, \overset{\circ}{0} \Big\| & m'_{1n} \cdots m'_{nn} \\ m_{1\,n-1} \cdots m_{n-1\,n-1} & \Big\| q, \overset{\circ}{0} \Big\| & m'_{1\,n-1} \cdots m'_{n-1\,n-1} \end{pmatrix}$$

is connected with both the 9-j symbols of U(n) and U($n-1$). A consequence of our result is the result obtained by Holman[9]: that the Weyl coefficients of U(n) are basically 6-j symbols of U($n-1$). This is easily understood when one considers the case of U(3). There a 9-j symbol in U(2) becomes a 6-j symbol when one of the terms (i.e., $p-q$) becomes zero. This is precisely what happens when one calculates the Weyl coefficients of U(3).

By putting another term (i.e., q) equal to zero in the 9-j symbol of U(n), we are able to evaluate a "stretched" 6-j symbol in U(n), which involves no sums. The generalized beta functions of Gel'fand and Graev[10]

for U(n) are found to be connected with the stretched 6-j symbols of U($n-1$) and the isoscalar factor of U($n-1$).

Another result is that the transpositional symmetry of the isoscalar factors, which is a consequence of the commutativity of upper and lower patterns of the boson polynomials, can now be interpreted as the transpositional symmetry of 9-j symbols, while the Regge symmetries of the 6-j symbol of U(2) can be interpreted as the symmetries of the Weyl coefficients of the double boson state of U(3) * U(3).

In Sec. 1 we shall show that the isoscalar factor

$$\begin{pmatrix} m_{1n} \cdots m_{n-1\,n} \; 0 & \Big\| p, \overset{\circ}{0} \Big\| & m'_{1n} \cdots m'_{n-1\,n} \; 0 \\ m_{1\,n-1} \cdots m_{n-1\,n-1} & \Big\| q, \overset{\circ}{0} \Big\| & m'_{1\,n-1} \cdots m'_{n-1\,n-1} \end{pmatrix}$$

in U(n) is connected with the doubly stretched 9-j symbol of U($n-1$) and that

$$\begin{pmatrix} m_{1n} \cdots m_{nn} & \Big\| p, \overset{\circ}{0} \Big\| & m'_{1n} \circ \circ m'_{nn} \\ m'_{1\,n-1} \circ \circ m'_{n-1\,n-1} & \Big\| q, \overset{\circ}{0} \Big\| & m'_{1\,n-1} \circ \circ m'_{n-1\,n-1} \end{pmatrix}$$

is connected with the 9-j symbols of both U(n) and U($n-1$).

In Sec. 2 we discuss a consequence of this result, i.e., the Weyl coefficients of U(n) are basically 6-j symbols in U($n-1$), a result first obtained by Holman.[9] With this result the transformation matrix $D^{[m]_n}_{(m')_{n-1}(m)_{n-1}}$ for U(n), either in terms of Euler angles or in terms of bosons a^j_i, can be written in a very simple form. Furthermore, we propose that the D matrix parametrized in terms of Euler angles should be diagonalized according to labels, denoted as z_i by us, which have simple properties under R conjugation, i.e., $z_i^* = -z_i$. In SU(3) these are just the isospin and hyper-

charge quantum numbers. Therefore, the D matrices diagonal in these quantum numbers will have direct physical applications. The D matrix of SU(3) diagonal in isospin and hypercharge obtained by Yeh[11] is explicitly given. Its relation with the boson polynomials is also given.

Another structural property that becomes clear in the double boson polynomial is that the Wigner coefficient of U(2) is contained in the double boson polynomial of U(2), and can therefore be written down by inspection. This is certainly the easiest way of deriving the Wigner coefficient of U(2).

In Sec. 3 we derive the expression for the "stretched" 6-j symbols of U(n), which involve no summations. The "stretched" 6-j symbols of U($n-1$) and the isoscalar factor in U($n-1$) are then found to be connected with the generalized beta functions of Gel'fand and Graev.

In the Appendix we give the phase relations between Wigner coefficients and 3-j symbols of U(n). This is by no means a trivial problem, since there is already disagreement in SU(3), when one considers that the phases of the "1-j" symbol in SU(3), obtained by de Swart[12] (or Resnikoff[5]), Ponzano,[4] and Baird and Biedenharn[13] are all different. Our principle is that in the case of U(2), the phase convention should be that of Eqs. (3.7.3), (3.7.5), and (3.7.6) of Edmonds,[14] a point on which almost everybody agrees. Then this phase should generalize to U(n). In doing so we find that one must use labels corresponding to j and m in SU(2), which have simple properties under R conjugation. These labels are denoted by $\epsilon_n \lambda_i$, and z_i, $i = 1, 2, \ldots, n-1$, by us. When these labels are used one finds that the phase convention of U(2) generalizes to U(n).

1. MULTIPLICITY-FREE WIGNER COEFFICIENTS OF U(n) AND 9-j SYMBOLS OF U(n-1)

To prepare ourselves for the desired result, we first define singly stretched 9-j symbols in U(n), following basically Sharp and von Baeyer,[15] who did it in the case of U(2). The 9-j symbol is defined by

$$
\sum_{\substack{(\alpha)_{n-1}(\beta)_{n-1} \\ (\gamma)_{n-1}(\delta)_{n-1} \\ (\theta)_{n-1}(\kappa)_{n-1}}}
\left\langle \begin{array}{c} [a]_n \\ (\alpha)_{n-1} \end{array} \right.
\left| \begin{array}{c} [b] \\ (\beta)_{n-1} \end{array} \right\rangle
\left\langle \begin{array}{cc} [a]_n & [b]_n \\ (\alpha)_{n-1} & (\beta)_{n-1} \end{array} \right|
\left. \begin{array}{c} [j]_n \\ (\theta)_{n-1} \end{array} \right\rangle
$$

$$
\times \left\langle \begin{array}{c} [c]_n \\ (\gamma)_{n-1} \end{array} \right.
\left| \begin{array}{c} [d]_n \\ (\delta)_{n-1} \end{array} \right\rangle
\left\langle \begin{array}{cc} [c]_n & [d]_n \\ (\gamma)_{n-1} & (\delta)_{n-1} \end{array} \right|
\left. \begin{array}{c} [k]_n \\ (\kappa)_{n-1} \end{array} \right\rangle
$$

$$
\times \left\langle \begin{array}{cc} [j]_n & [k]_n \\ (\theta)_{n-1} & (\kappa)_{n-1} \end{array} \right|
\left. \begin{array}{c} [e]_n \\ (e)_{\max} \end{array} \right\rangle
$$

$$
= \sum_{\substack{(\alpha)_{n-1}(\beta)_{n-1} \\ (\delta)_{n-1}(\gamma)_{n-1}(\theta)_{n-1} \\ (\epsilon)_{n-1}(\phi)_{n-1}}}
\left\langle \begin{array}{c} [a]_n \\ (\alpha)_{n-1} \end{array} \right.
\left| \begin{array}{c} [c]_n \\ (\gamma)_{n-1} \end{array} \right\rangle
$$

$$
[f]_n[g]_n
$$

$$
\times \left\langle \begin{array}{cc} [a]_n & [c]_n \\ (\alpha)_{n-1} & (\gamma)_{n-1} \end{array} \right|
\left. \begin{array}{c} [f]_n \\ (\epsilon)_{n-1} \end{array} \right\rangle
\left| \begin{array}{c} [b]_n \\ (\beta)_{n-1} \end{array} \right\rangle
$$

$$
\times \left| \begin{array}{c} [d]_n \\ (\delta)_{n-1} \end{array} \right\rangle
\left\langle \begin{array}{cc} [b]_n & [d]_n \\ (\beta)_{n-1} & (\delta)_{n-1} \end{array} \right|
\left. \begin{array}{c} [g]_n \\ (\phi)_{n-1} \end{array} \right\rangle
$$

$$
\times \left\langle \begin{array}{cc} [f]_n & [g]_n \\ (\epsilon)_{n-1} & (\phi)_{n-1} \end{array} \right|
\left. \begin{array}{c} [e]_n \\ (e)_{\max} \end{array} \right\rangle
X \begin{vmatrix} [a]_n & [b]_n & [j]_n \\ [c]_n & [d]_n & [k]_n \\ [f]_n & [g]_n & [e]_n \end{vmatrix}
$$

$$
\times (\dim|j]_n \dim[k]_n \dim[f]_n \dim[g]_n)^{1/2} \tag{1.1}
$$

where a square bracket means a row of numbers, thus, e.g., $[f]_n = (f_{1n}, f_{2n}, \ldots, f_{nn})$, while parenthesis means a Gel'fand pattern. Throughout this paper we are dealing with the multiplicity-free Wigner coefficients of U(n); therefore, the Wigner coefficients need not carry a multiplicity label, or an upper operator pattern. Max means that the state is in its maximum weight. Thus, e.g., for U(2)

$$
\left| \begin{array}{c} [e]_n \\ (e)_{\max} \end{array} \right\rangle = \left| \begin{array}{cc} e_{12} & e_{22} \\ & e_{12} \end{array} \right\rangle.
$$

Take the scalar product of (1.1) with

$$
\left| \begin{array}{c} \dot{a}, 0 \\ (\dot{a})_{n-1} \\ \max \end{array} \right\rangle
\left| \begin{array}{c} \dot{c}, 0 \\ (\dot{c})_{n-1} \\ \max \end{array} \right\rangle,
$$

the left side is

$$
\sum_{\substack{(\beta)_{n-1} \\ (\delta)_{n-1}(\theta)_{n-1}}}
\left| \begin{array}{c} [b]_n \\ (\beta)_{n-1} \end{array} \right\rangle
\left\langle \begin{array}{cc} \dot{a}, 0 & [b]_n \\ (\dot{a})_{n-1} & (\beta)_{n-1} \\ \max \end{array} \right|
\left. \begin{array}{c} [j]_n \\ (\theta)_{n-1} \end{array} \right\rangle
\left| \begin{array}{c} [d]_n \\ (\delta)_{n-1} \end{array} \right\rangle
$$

$$
\left\langle \begin{array}{cc} \dot{c}, 0 & [d]_n \\ (\dot{c})_{n-1} & (\delta)_{n-1} \\ \max \end{array} \right|
\left. \begin{array}{c} [k]_n \\ (e)_{n-1} - (\theta)_{n-1} \\ \max \end{array} \right\rangle
$$

$$
\times \left\langle \begin{array}{cc} [j]_n & [k]_n \\ (\theta)_{n-1} & (e)_{n-1} - (\theta)_{n-1} \\ \max \end{array} \right|
\left. \begin{array}{c} [e]_n \\ (e)_{n-1} \\ \max \end{array} \right\rangle. \tag{1.2}
$$

Since the state

$$
\left| \begin{array}{c} \dot{a}, 0 \\ (\dot{a})_{n-1} \\ \max \end{array} \right\rangle
$$

is the conjugate state of

$$
\left| \begin{array}{c} a, \dot{0} \\ \dot{0} \end{array} \right\rangle,
$$

we can write the Wigner coefficient

$$
\left\langle \begin{array}{cc} \dot{a}, 0 & [b]_n \\ (\dot{a})_{n-1}, & (\beta)_{n-1} \\ \max \end{array} \right|
\left. \begin{array}{c} [j]_n \\ (\theta)_{n-1} \end{array} \right\rangle
$$

$$
= (-1)^y \left(\frac{\dim[j]_n}{\dim[b]_n} \right)^{1/2}
\left\langle \begin{array}{cc} [j]_n & a, \dot{0} \\ (\theta)_{n-1}, & \dot{0} \end{array} \right|
\left. \begin{array}{c} [b]_n \\ (\theta)_{n-1} \end{array} \right\rangle, \tag{1.3}
$$

where y is some phase factor. In the case of U(2), for $\sum[j]_n + a = \sum[b]_n$, $(-1)^y = 1$. We have obtained the phase factors for U(n). The details are given in the Appendix. There it is shown that y is a function of $[b]_n$ only, if nonvanishing. Similarly we have

$$\left\langle \begin{matrix} \dot{c}, & 0 & [d]_n \\ (\dot{c})_{n-1}, & (\delta)_{n-1} \\ \text{max} & \text{max} \end{matrix} \middle| \begin{matrix} [k]_n \\ (e)_{n-1} - (\theta)_{n-1} \end{matrix} \right\rangle$$

$$= (-1)^{y(d)} \left(\frac{\dim[k]_n}{\dim[d]_n} \right)^{1/2} \left\langle \begin{matrix} [k]_n & c, \dot{0} \\ (e)_{n-1} - (\theta)_{n-1}, \dot{0} & \\ \text{max} & \end{matrix} \middle| \begin{matrix} [d]_n \\ (e)_{n-1} - (\theta)_{n-1} \\ \text{max} \end{matrix} \right\rangle \tag{1.4}$$

On the right the only value of $[f]_n$, limited to the conjugate of the totally symmetric representation $[f, \dot{0}]$ $+[f]_n = [a \dot{+} c, 0]_n$. This limitation ensures that the Wigner coefficients in (1.1) are multiplicity-free. Thus the right side becomes

$$\sum_{\substack{(\theta)_{n-1} \\ [g]_n}} \left| \begin{matrix} [b]_n \\ (\theta)_{n-1} \end{matrix} \right\rangle \left| \begin{matrix} [d]_n \\ (e)_{n-1} - (\theta)_{n-1} \\ \text{max} \end{matrix} \right\rangle$$

$$\times \left\langle \begin{matrix} [b]_n & [d]_n & [g]_n \\ (\theta)_{n-1}, (e)_{n-1} - (\theta)_{n-1} & (e)_{n-1} \\ \text{max} & \text{max} \end{matrix} \right\rangle$$

Now take the scalar product of both sides with

$$\sum_{(\theta)_{n-1}} \left| \begin{matrix} [b]_n \\ (\theta)_{n-1} \end{matrix} \right\rangle \left| \begin{matrix} [d]_n \\ (e)_{n-1} - (\theta)_{n-1} \\ \text{max} \end{matrix} \right\rangle \left\langle \begin{matrix} [b]_n & [d]_n & [g]_n \\ (\theta)_{n-1}, (e)_{n-1} - (\theta)_{n-1} & (e)_{n-1} \\ \text{max} & \text{max} \end{matrix} \right\rangle;$$

we obtain

$$X \left| \begin{matrix} [\dot{a}, 0]_n & [b]_n [j]_n \\ [\dot{c}, \dot{0}]_n & [d]_n [k]_n \\ [a \dot{+} c, 0]_n & [g]_n [e]_n \end{matrix} \right| = \sum_{(\theta)_{n-1}} \left\langle \begin{matrix} [b]_n & [d]_n & [g]_n \\ (\theta)_{n-1}, (e)_{n-1} - (\theta)_{n-1} & (e)_{n-1} \\ \text{max} & \text{max} \end{matrix} \right\rangle \left\langle \begin{matrix} [j]_n & [a, \dot{0}]_n & [b]_n \\ (\theta)_{n-1} & \dot{0} & (\theta)_{n-1} \end{matrix} \right\rangle \left\langle \begin{matrix} [k]_n & [c, \dot{0}]_n & [d]_n \\ (e)_{n-1} - (\theta)_{n-1}, & \dot{0} & (e)_{n-1} - (\theta)_{n-1} \\ \text{max} & & \text{max} \end{matrix} \right\rangle$$

$$\times \left\langle \begin{matrix} [j]_n & [k]_n & [e]_n \\ (\theta)_{n-1} & (e)_{n-1} - (\theta)_{n-1} & (e)_{n-1} \\ & \text{max} & \text{max} \end{matrix} \right\rangle \left\langle \begin{matrix} [e]_n & [a + c, \dot{0}]_n & [g]_n \\ (e)_{n-1} & \dot{0} & (e)_{n-1} \\ \text{max} & & \text{max} \end{matrix} \right\rangle^{-1}$$

$$\times (\dim[a + c, \dot{0}]_n \dim[b]_n \dim[d]_n \dim[e]_n)^{-1/2} (-1)^{y(d) + y(b) + y(g)}. \tag{1.7}$$

To obtain the result for U(n), let us first consider the case for U(3). The scalar product of the boson polynomials

$$\left\langle \begin{matrix} 0 \end{matrix} \middle| \begin{pmatrix} m'_{13} \\ m'_{13} \, m'_{23} \\ m'_{13} \, m'_{23} \, 0 \\ m'_{12} \, m'_{22} \\ m'_{12} \end{pmatrix}^{\dagger} B \begin{pmatrix} 0 \\ p \, 0 \\ p \, 0 \, 0 \\ q \, 0 \\ 0 \end{pmatrix}^{\dagger} \begin{pmatrix} m'_{13} \\ m_{13} \, m_{23} \\ m_{13} \, m_{23} \, 0 \\ m_{12} \, m_{22} \\ m'_{12} \end{pmatrix} \middle| 0 \right\rangle \tag{1.8}$$

can obviously be evaluated in terms of the bosons in U(2)×U(2) only, since the irreducible rep. labels ($m_{13} \, m_{23} \, 0$), ($p \, 0 \, 0$) and ($m'_{13} \, m'_{23} \, 0$) can all be considered as tensors transforming within U(2)×U(2) entirely. Using the rules for the coupling of bosons,[16] we find that the scalar product in (1.8) is equal to

$$\sum_{p_3^{2'}} \left\langle \begin{pmatrix} m'_{12} + m'_{22} - m'_{23} + p_3^{2'} \\ m_{12} \, m_{22} \\ m_{12} \, m_{22} \, 0 \\ m_{12} \, m_{22} \\ m'_{12} \end{pmatrix} B \begin{pmatrix} 0 \\ q \, 0 \\ q \, 0 \, 0 \\ q \, 0 \\ 0 \end{pmatrix} \middle| \begin{matrix} m'_{12} + m'_{22} - m'_{23} + p_3^{2'} \\ m'_{12} \, m'_{22} \\ m'_{12} \, m'_{22} \, 0 \\ m'_{12} \, m'_{22} \\ m'_{12} \end{matrix} \right\rangle \left[\frac{M(m_{12} m_{22})}{M(m'_{12} m'_{22})} \right]^{1/2}$$

$$\times \left\langle \begin{matrix} [a \dot{+} c, 0]_n [g]_n & [e]_n \\ (a \dot{+} c)_{n-1}, & (e)_{n-1} & (e)_{n-1} \\ \text{max} & \text{max} & \text{max} \end{matrix} \right\rangle$$

$$\times X \left| \begin{matrix} [\dot{a}, 0]_n & [b]_n [j]_n \\ [\dot{c}, 0]_n & [d]_n [k]_n \\ [a \dot{+} c, 0]_n & [g]_n [e]_n \end{matrix} \right|$$

$$\times (\dim[j]_n \dim[k]_n \dim[g]_n \dim[a + c, \dot{0}]_n)^{1/2}. \tag{1.5}$$

Again we can write

$$\left\langle \begin{matrix} [a \dot{+} c, 0]_n & [g]_n & [e]_n \\ (a \dot{+} c)_{n-1} & (e)_{n-1} & (e)_{n-1} \\ \text{max} & \text{max} & \text{max} \end{matrix} \right\rangle$$

$$= (-1)^{y(g)} \left(\frac{\dim[e]_n}{\dim[g]_n} \right)^{1/2} \left\langle \begin{matrix} [e]_n & [a + c, \dot{0}]_n & [g]_n \\ (e)_{n-1}, & \dot{0} & (e)_{n-1} \\ \text{max} & & \text{max} \end{matrix} \right\rangle. \tag{1.6}$$

$$\times \left\langle \begin{matrix} W_3' - p_3^{2\prime} \\ W_3 \; 0 \\ W_3 \; 0 \; 0 \\ 0 \; 0 \\ 0 \end{matrix} \middle| B \begin{pmatrix} 0 \\ p-q \; 0 \\ p-q \; 0 \; 0 \\ 0 \; 0 \\ 0 \end{pmatrix} \middle| \begin{matrix} W_3' - p_3^{2\prime} \\ W_3' \; 0 \\ W_3' \; 0 \; 0 \\ 0 \; 0 \\ 0 \end{matrix} \right\rangle \left[\frac{W_3!}{W_3'!}\right]^{1/2} \frac{p!}{(p-q)!\,q!} \left\langle \begin{matrix} 0 \\ p \; 0 \\ p \; 0 \; 0 \\ q \; 0 \\ 0 \end{matrix} \middle| \begin{matrix} 0 \\ q \; 0 \\ q \; 0 \; 0 \\ p \; 0 \\ 0 \end{matrix} \middle| \begin{matrix} 0 \\ p-q \; 0 \\ p-q \; 0 \; 0 \\ 0 \; 0 \\ 0 \end{matrix} \right\rangle$$

$$\times \left\langle \begin{matrix} m_{13}' \\ m_{13}' \; m_{23}' \\ m_{13}' \; m_{23}' \; 0 \\ m_{12}' \; m_{22}' \\ m_{12}' \end{matrix} \middle| \begin{matrix} W_3' - p_3^{2\prime} \\ W_3' \; 0 \\ W_3' \; 0 \; 0 \\ 0 \; 0 \\ 0 \end{matrix} \middle| \begin{matrix} m_{13}' - W_3' + p_3^{2\prime} \\ m_{12}' \; m_{22}' \\ m_{12}' \; m_{22}' \; 0 \\ m_{12}' \; m_{22}' \\ m_{12}' \end{matrix} \right\rangle \left\langle \begin{matrix} \frac{1}{2}p & \frac{1}{2}(m_{13}-m_{23}) \\ \frac{1}{2}p, & m_{13}' - \frac{1}{2}(m_{13}+m_{23}) \end{matrix} \middle| \begin{matrix} \frac{1}{2}(m_{13}'-m_{23}') \\ \frac{1}{2}(m_{13}'-m_{23}') \end{matrix} \right\rangle \left[\frac{\mathcal{M}(m_{13}'m_{23}')}{W_3'!\mathcal{M}(m_{12}'m_{22}')}\right]^{1/2} \left[\frac{\mathcal{M}(m_{13}m_{23})}{W_3!\mathcal{M}(m_{12}m_{22})}\right]^{1/2}$$

$$\times \left\langle \begin{matrix} m_{13}' \\ m_{13} \; m_{23} \\ m_{13} \; m_{23} \; 0 \\ m_{12} \; m_{22} \\ m_{12}' \end{matrix} \middle| \begin{matrix} W_3' - p_3^{2\prime} \\ W_3 \; 0 \\ W_3 \; 0 \; 0 \\ 0 \; 0 \\ 0 \end{matrix} \middle| \begin{matrix} m_{12}' + m_{22}' - m_{23}' + p_3^{2\prime} \\ m_{12} \; m_{22} \\ m_{12} \; m_{22} \; 0 \\ m_{12} \; m_{22} \\ m_{12}' \end{matrix} \right\rangle , \tag{1.9}$$

where

$$W_3 = m_{13} + m_{23} - m_{12} - m_{22}, \quad W_3' = m_{13}' + m_{23}' - m_{12}' - m_{22}', \quad W_3' + p - q = W_3.$$

$\mathcal{M}$ is the measure

$$\mathcal{M} = \prod_{i=1}^{n} (m_{in} + n - i)! / \prod_{i<j}^{n} (m_{in} - m_{jn} + j - i) \tag{1.10}$$

Using (1.7), we find (1.9) is equal to

$$\left\langle \begin{matrix} m_{12} \; m_{22} \\ m_{12}' \end{matrix} \middle| \begin{matrix} q \; 0 \\ 0 \end{matrix} \middle| \begin{matrix} m_{12}' \; m_{22}' \\ m_{12}' \end{matrix} \right\rangle \left[\frac{p!}{(p-q)!\,q!} (W_3+1)(m_{12}-m_{22}+1)(p+1)(m_{13}-m_{23}+1)\right]^{1/2}$$

$$\times \left[\frac{W_3!}{W_3'!} \frac{\mathcal{M}(m_{12}m_{22})}{\mathcal{M}(m_{12}'m_{22}')}\right]^{1/2} X \begin{vmatrix} \frac{1}{2}(p-q) & \frac{1}{2}W_3 & \frac{1}{2}W_3' \\ \frac{1}{2}q & \frac{1}{2}(m_{12}-m_{22}) & \frac{1}{2}(m_{12}'-m_{22}') \\ \frac{1}{2}p & \frac{1}{2}(m_{13}-m_{23}) & \frac{1}{2}(m_{13}'-m_{23}') \end{vmatrix} . \tag{1.11}$$

On the other hand, the scalar product in (1.8), according to the factorization lemma,[17] is also equal to

$$\left[\frac{\mathcal{M}(m_{13}m_{23})}{\mathcal{M}(m_{13}'m_{23}')}\right]^{1/2} \left\langle \begin{matrix} m_{13} \; m_{23} \; 0 \\ m_{12} \; m_{22} \end{matrix} \middle| \begin{matrix} p \; 0 \; 0 \\ q \; 0 \end{matrix} \middle| \begin{matrix} m_{13}' \; m_{23}' \; 0 \\ m_{12}' \; m_{22}' \end{matrix} \right\rangle \left\langle \begin{matrix} m_{12} \; m_{22} \\ m_{12}' \end{matrix} \middle| \begin{matrix} q \; 0 \\ 0 \end{matrix} \middle| \begin{matrix} m_{12}' \; m_{22}' \\ m_{12}' \end{matrix} \right\rangle \left\langle \begin{matrix} m_{13} \; m_{23} \\ m_{13}' \end{matrix} \middle| \begin{matrix} p \; 0 \\ 0 \end{matrix} \middle| \begin{matrix} m_{13}' \; m_{23}' \\ m_{13}' \end{matrix} \right\rangle . \tag{1.12}$$

Equating (1.11) and (1.12), we obtain

$$\left\langle \begin{matrix} m_{13} \; m_{23} \; 0 \\ m_{12} \; m_{22} \end{matrix} \middle| \begin{matrix} p \; 0 \; 0 \\ q \; 0 \end{matrix} \middle| \begin{matrix} m_{13}' \; m_{23}' \; 0 \\ m_{12}' \; m_{22}' \end{matrix} \right\rangle = \left[\frac{\mathcal{M}(m_{13}' \; m_{23}')W_3!\mathcal{M}(m_{12} \; m_{22})}{\mathcal{M}(m_{13} \; m_{23})W_3'!\mathcal{M}(m_{12}'m_{22}')}\right]^{1/2} \left[\frac{p!}{(p-q)!\,q!}\right]^{1/2}$$

$$\times [(m_{13}' - m_{23}' + 1)(m_{12} - m_{22} + 1)(W_3 + 1)(p+1)]^{1/2} X \begin{vmatrix} \frac{1}{2}q & \frac{1}{2}(m_{12} - m_{22}) & \frac{1}{2}(m_{12}' - m_{22}') \\ \frac{1}{2}(p-q) & \frac{1}{2}W_3 & \frac{1}{2}W_3' \\ \frac{1}{2}p & \frac{1}{2}(m_{13} - m_{23}) & \frac{1}{2}(m_{13}' - m_{23}') \end{vmatrix} . \tag{1.13}$$

This result agrees with Ališauskas[6] and Holman.[9] Let us also note that in the case of U(2) the analogous result of Eq. (1.13) becomes

$$\left\langle \begin{matrix} 2j_1 + 2j_2, \; 0 \\ j_1 + j_2 + m_1 + m_2 \end{matrix} \middle| \begin{matrix} 2j_2, \; 0 \\ j_2 + m_2 \end{matrix} \middle| \begin{matrix} 2j_1, \; 0 \\ j_1 + m_1 \end{matrix} \right\rangle = \left[\frac{\mathcal{M}(2j_1)\mathcal{M}(j_1 + j_2 - m_1 - m_2)\mathcal{M}(j_1 + j_2 + m_1 + m_2)}{\mathcal{M}(2j_1 + 2j_2)\mathcal{M}(j_1 - m_1)\mathcal{M}(j_1 + m_1)}\right]^{1/2}$$

$$\times \left[\frac{M(2j_2)}{M(j_2-m_2)M(j_2+m_2)}\right]^{1/2} X \begin{vmatrix} \frac{1}{2}(j_2+m_2) & \frac{1}{2}(j_1+j_2+m_1+m_2) & \frac{1}{2}(j_1+m_1) \\ \frac{1}{2}(j_2-m_2) & \frac{1}{2}(j_1+j_2-m_1-m_2) & \frac{1}{2}(j_1-m_1) \\ j_2 & j_1+j_2 & j_1 \end{vmatrix} \tag{1.14}$$

where the 9-j symbol is in U(1), and is of course, equal to 1. Equation (1.14) agrees with the result in U(2), i.e.,

$$C^{j_1\,j_2\,j_1+j_2}_{m_1\,m_2\,m_1+m_2} = \left[\frac{(2j_1)!\,(2j_2)!\,(j_1+j_2+m_1+m_2)!\,(j_1+j_2-m_1-m_2)!}{(2j_1+2j_2)!\,(j_1-m_1)!\,(j_1+m_1)!\,(j_2-m_2)!\,(j_2+m_2)!}\right]^{1/2} \tag{1.15}$$

The argument used for U(3) can undoubtedly be extended to U(n). Thus with the 9-j symbol defined by (1.7) and with the mapping

$$\begin{pmatrix} a, & \overset{\circ}{0} \\ & \overset{\circ}{0} \end{pmatrix} = \begin{pmatrix} \overset{\circ}{a}, & 0 \\ & \overset{\circ}{a} \end{pmatrix}^* \rightarrow \begin{pmatrix} p-q, & \overset{\circ}{0} \\ & 0 \end{pmatrix}, \quad \begin{pmatrix} c, & \overset{\circ}{0} \\ & \overset{\circ}{0} \end{pmatrix} = \begin{pmatrix} \overset{\circ}{c}, & 0 \\ & \overset{\circ}{c} \end{pmatrix}^* \rightarrow \begin{pmatrix} q, & \overset{\circ}{0} \\ & 0 \end{pmatrix},$$

$$\begin{pmatrix} [b]_n \\ (\theta)_{n-1} \end{pmatrix} \rightarrow \begin{pmatrix} W_n, & \overset{\circ}{0} \\ (\theta)_{n-1} \end{pmatrix} \quad \begin{pmatrix} j_n \\ (\theta)_{n-1} \end{pmatrix} \rightarrow \begin{pmatrix} W'_n, & \overset{\circ}{0} \\ (\theta)_{n-1} \end{pmatrix}, \quad \begin{pmatrix} [d]_n \\ (e)_{n-1} - (\theta)_{n-1} \\ \text{max} \end{pmatrix} \rightarrow \begin{pmatrix} m_{1\,n-1}\cdots m_{n-1\,n-1}, 0 \\ (e)_{n-1} - (\theta)_{n-1} \\ \text{max} \end{pmatrix} \quad \begin{pmatrix} [k]_n \\ (e)_{n-1} - (\theta)_{n-1} \\ \text{max} \end{pmatrix}$$

$$\rightarrow \begin{pmatrix} m'_{1\,n-1}\cdots m'_{n-1\,n-1},0 \\ (e)_{n-1} - (\theta)_{n-1} \\ \text{max} \end{pmatrix} \quad \begin{pmatrix} [g]_n \\ (e)_{n-1} \\ \text{max} \end{pmatrix} \rightarrow \begin{pmatrix} m_{1n}\cdots m_{n-1\,n},0 \\ (e)_{n-1} \\ \text{max} \end{pmatrix}, \quad \begin{pmatrix} [e]_n \\ (e)_{\text{max}} \end{pmatrix} \rightarrow \begin{pmatrix} m'_{1n}\cdots m'_{n-1\,n},0 \\ \text{max} \end{pmatrix}, \quad (\theta)_{n-1} = \begin{pmatrix} W'_n - p_n^{n-1\prime} & \overset{\circ}{0} \\ W'_n - p_n^{n-1\prime} - p_n^{n-2\prime} & \overset{\circ}{0} \\ \cdots \\ W'_n - p_n^{n-1\prime} \cdots - p_n^{2\prime} \end{pmatrix},$$

$$\tag{1.16}$$

$\sum_{(\theta)_{n-1}} = \sum_p p_n^{i\prime}$, $i = 2, 3, \ldots, n-1$. We obtain

$$\left\langle \begin{matrix} m_{1n}\cdots m_{n-1\,n}, & 0 \\ m_{1\,n-1}\cdots m_{n-1\,n-1} \end{matrix} \middle\| \begin{matrix} p, & \overset{\circ}{0} \\ q, & \overset{\circ}{0} \end{matrix} \middle\| \begin{matrix} m'_{1n}\cdots m'_{n-1\,n}, & 0 \\ m'_{1\,n-1}\cdots m'_{n-1\,n-1} \end{matrix} \right\rangle = \left[\frac{M(m'_{1n}\cdots m'_{n-1\,n})W_n!}{M(m_{1n}\cdots m_{n-1\,n})W'_n!}\right]^{1/2} \left[\frac{M(m_{1\,n-1}\cdots m_{n-1\,n-1})p!}{M(m'_{1\,n-1}\cdots m_{n-1\,n-1})(p-q)!\,q!}\right.$$

$$\times \dim(m'_{1\,n-1}\cdots m'_{n-1\,n-1})\dim(m_{1\,n-1}\cdots m_{n-1\,n-1}) \Big]^{1/2} [\dim[W_n,\overset{\circ}{0}]_{n-1}\dim[p,\overset{\circ}{0}]_{n-1}]^{1/2}$$

$$\times (-1)^{y(m_{1\,n-1}\cdots m_{n-1\,n-1})+y(W_n)+y(m_{1\,n}\cdots m_{n-1\,n})} \times X \begin{vmatrix} [\overset{\circ}{q}, 0]_{n-1} & [m_{1\,n-1}\cdots m_{n-1\,n-1}] & [m'_{1\,n-1}\cdots m'_{n-1\,n-1}] \\ [p - q, 0]_{n-1} & [W_n, \overset{\circ}{0}]_{n-1} & [W'_n, \overset{\circ}{0}] \\ [\overset{\circ}{p}, 0]_{n-1} & [m_{1\,n}\cdots m_{n-1\,n}] & [m'_{1n}\cdots m'_{n-1\,n-1}] \end{vmatrix}. \tag{1.17}$$

Though the 9-j symbol defined in(1.7) is singly stretched, the 9-j symbol we obtain in Eq. (1.17) is doubly stretched, since $W'_n + p - q = W_n$.

The isoscalar factor in (1.17) does not exhaust all possibilities of multiplicity-free Wigner coefficients of U(n). The most general multiplicity-free Wigner coefficient is

$$\left\langle \begin{matrix} m_{1n}\cdots m_{nn} \\ m_{1\,n-1}\cdots m_{n-1\,n-1} \end{matrix} \middle\| \begin{matrix} p, & \overset{\circ}{0} \\ q, & \overset{\circ}{0} \end{matrix} \middle\| \begin{matrix} m'_{1n}\cdots m'_{nn} \\ m'_{1\,n-1}\cdots m'_{n-1\,n-1} \end{matrix} \right\rangle. \tag{1.18}$$

Without loss of generality one can rewrite the state on the right so that $m'_{nn} = 0$. However, m_{nn} does not have to be zero. In this case the isoscalar factor will truly depend on the 9-j symbol of U(n) as well. Proceeding in the same fashion as before, we obtain the following equation for U(3):

$$\left\langle \begin{matrix} m_{13}\,m_{23}\,m_{33} \\ m_{12}\,m_{22} \end{matrix} \middle\| \begin{matrix} p\,0\,0 \\ q\,0 \end{matrix} \middle\| \begin{matrix} m'_{13}\,m'_{23}\,0 \\ m'_{12}\,m'_{22} \end{matrix} \right\rangle X \begin{vmatrix} [p-q,0,0] & [W'_3,0,0] & [W_3,0,0] \\ [q,0,0] & [m'_{12},m'_{22},0] & [m_{12},m_{22},0] \\ [p,0,0] & [m'_{13},m'_{23},0] & [m_{13},m_{23},m_{33}] \end{vmatrix}$$

$$= \left[\frac{M(m_{13}m_{23}m_{33})}{M(m'_{13}m'_{23})}\right]^{1/2} \left\langle \begin{matrix} m_{13}\,m_{23}\,m_{33} \\ m_{13}\,m_{23} \end{matrix} \middle\| \begin{matrix} p\,0\,0 \\ p\,0 \end{matrix} \middle\| \begin{matrix} m'_{13}\,m'_{23}\,0 \\ m'_{13}\,m'_{23} - m_{33} \end{matrix} \right\rangle^{-1} \left[\frac{p!\,M(m_{12}m_{22})}{(p-q)!\,q!\,M(m'_{12}m'_{22})}\right]^{1/2}$$

$$\times \sum_{p_3^{3\prime}} \left\langle \begin{matrix} W_3\,0\,0 \\ W_3 - p_3^{3\prime}\,0 \end{matrix} \middle\| \begin{matrix} p-q\,0\,0 \\ p-q\,0 \end{matrix} \middle\| \begin{matrix} W'_3\,0\,0 \\ W'_3 - p_3^{3\prime}\,0 \end{matrix} \right\rangle \left\langle \begin{matrix} m_{12}\,m_{22}\,0 \\ m_{12}\,m_{22} + p_3^{3\prime} - m_{33} \end{matrix} \middle\| \begin{matrix} q\,0\,0 \\ q\,0 \end{matrix} \middle\| \begin{matrix} m'_{12}\,m'_{22}\,0 \\ m'_{12}\,m'_{22} + p_3^{3\prime} - m_{33} \end{matrix} \right\rangle$$

$$\times \left\langle \begin{matrix} m_{13} & m_{23} & m_{33} \\ m_{13} & m_{23} \end{matrix} \middle\| \begin{matrix} W_3 & 0 & 0 \\ W_3 - p_3^{3\prime} & 0 \end{matrix} \middle\| \begin{matrix} m_{12} & m_{22} & 0 \\ m_{12} & m_{22} + p_3^{3\prime} - m_{33} \end{matrix} \right\rangle [(W_3 - p_3^{3\prime} + 1)(m_{13}' - m_{23}' + 1)]^{1/2}$$

$$\times [(m_{12} - m_{22} + m_{33} - p_3^{3\prime} + 1)(p + 1)]^{1/2} X \begin{vmatrix} \frac{1}{2}q & \frac{1}{2}(m_{12} - m_{22} + m_{33} - p_3^{3\prime}) & \frac{1}{2}(m_{12}' - m_{22}' - p_3^{3\prime} + m_{33}) \\ \frac{1}{2}(p - q) & \frac{1}{2}(W_3 - p_3^{3\prime}) & \frac{1}{2}(W_3' - p_3^{3\prime}) \\ \frac{1}{2}p & \frac{1}{2}(m_{13} - m_{23}) & \frac{1}{2}(m_{13}' - m_{23}' + m_{33}) \end{vmatrix} \qquad (1.19)$$

The isoscalar factor on the left side of (1.19) has been evaluated by Chacón et al.,[7] Ališauskas et al.[8] All the other isoscalar factors in (1.19) do not involve any sums. Thus we can use (1.19) to evaluate the 9-j symbol of U(3) on the left of Eq. (1.19). In the case of U(3), this 9-j symbol, like the isoscalar factor, contains a sum over two indices. In U(n) the 9-j symbol will contain a sum over ($n - 1$) indices.

Jucys[18] and Holman[9] have obtained transpositional symmetry for the isoscalar factor in (1.19). In view of (1.17) and (1.19) this symmetry, when applied to the 9-j symbols, just means the transpositional symmetry of 9-j symbols between rows and columns. However, as is pointed out by Bincer[19] and Louck and Biedenharn,[16c] the transpositional symmetry of U(2)$_*$ U(2) gives the new Regge symmetry[20] for the 3-j symbol of U(2), which goes beyond the "classical symmetry" of the 3-j symbols. In Sec. 4 we shall discuss the Regge symmetry[21] of the 6-j symbol of U(2).

2. CONSEQUENCES OF THE RESULT

It has been shown by Holman[9] and Louck and Biedenharn[16] that the boson polynomial in U(n)$_*$ U(n) is connected with the isoscalar factor of the totally symmetric representation $(p, \overset{\circ}{0})$. Thus

$$\mathcal{M}(m)^{-1/2} B \begin{pmatrix} (m') \\ [m] \\ (m) \end{pmatrix} (A_n)$$

$$= \sum_{\substack{[\mu]_{n-1}(b)(\mu) \\ (b')(\mu')}} \left\langle \begin{matrix} [m] \\ (m) \end{matrix} \middle| \begin{matrix} [b, \overset{\circ}{0}] \\ (b) \end{matrix} \middle| \begin{matrix} [m']_{n-1} & 0 \\ (\mu) \end{matrix} \right\rangle$$

$$\times \left\langle \begin{matrix} [m']_{n-1} \\ (m') \end{matrix} \middle| \begin{matrix} [b', \overset{\circ}{0}] \\ (b') \end{matrix} \middle| \begin{matrix} [\mu]_{n-1} \\ (\mu') \end{matrix} \right\rangle [b! \, b'! / \mathcal{M} [\mu]_{n-1}]^{-1/2}$$

$$\times B \begin{pmatrix} (0) \\ [b, \overset{\circ}{0}] \\ (b) \end{pmatrix} (A_n) B \begin{pmatrix} (b') \\ \overset{(max)}{[b, \overset{\circ}{0}]} \\ (0) \end{pmatrix} (A_n) B \begin{pmatrix} (\mu') \\ [\mu]_{n-1} \\ (\mu) \end{pmatrix} (A_{n-1}).$$

$$(2.1)$$

This time, without loss of generality, we can put $m_{nn} = 0$. Then Eq. (1.17) can be applied to the first term on the right-hand side of (2.1).

The Weyl coefficients of U(n),

$$\left\langle \begin{matrix} [m]_n \\ [m]_{n-1} \end{matrix} \middle| W_{n-1\,n} \middle| \begin{matrix} [m]_n \\ [m']_{n-1} \end{matrix} \right\rangle$$

can be most easily evaluated by equating the boson A_n to

$$A_n = (a_i^j) = \begin{pmatrix} 1 & & & & & \\ & 1 & & & 0 & \\ & & \cdot & & & \\ & & & \cdot & & \\ & & & & \cdot & \\ & & & & & 1 \\ 0 & & & & 0 & 1 \\ & & & & & 1 & 0 \end{pmatrix}.$$

Thus the bosons $a_i^i = 1$, for $i = 1, 2, \ldots, n - 2$, $a_{n-1}^n = 1$, $a_n^{n-1} = 1$ and all other bosons in Eq. (2.1) must be equated to zero. This gives the constraints:

$$W_n = W'_{n-1} \quad \text{and} \quad W_m = W'_m \quad \text{for } m < n - 1. \qquad (2.2)$$

Then (2.1) immediately gives

$$[\mathcal{M}(m)]^{-1/2} B \begin{pmatrix} (m') \\ [m] \\ (m) \end{pmatrix} (A_n)$$

$$= \left\langle \begin{matrix} m_{1n} \circ \circ \circ m_{n-1\,n} & 0 \\ [m]_{n-1} \end{matrix} \middle\| \begin{matrix} b, \overset{\circ}{0} \\ W_{n-1}, 0 \end{matrix} \middle\|_{\circ} \begin{matrix} m'_{1\,n-1} \circ \circ \circ m'_{n-1\,n-1} & 0 \\ [m]_{n-2} \end{matrix} \right\rangle$$

$$\times \delta_{W_n, W'_{n-1}} \delta_{(m)_{n-2}(m')_{n-2}}$$

$$\times \left\langle \begin{matrix} m'_{1\,n-1} \circ \circ \circ m'_{n-1\,n-1} & | W'_{n-1}, & \overset{\circ}{0} \\ (m)_{n-2} & | W_{n-2}, & \overset{\circ}{0} \end{matrix} \middle| \begin{matrix} m_{1\,n-2} \circ \circ \circ m_{n-2\,n-2} & 0 \\ (m)_{n-3} \end{matrix} \right\rangle$$

$$\times [b! \, W'_{n-1}! / \mathcal{M}[m]_{n-2}]^{-1/2} \left\langle \begin{matrix} [m]_{n-1} \\ (m)_{n-2} \end{matrix} \middle| \begin{matrix} b, & \overset{\circ}{0} \\ W_{n-2}, 0 \end{matrix} \middle| \begin{matrix} [m]_{n-2} \\ (m)_{n-3} \end{matrix} \right\rangle.$$

$$(2.3)$$

Since the Weyl coefficient is independent of $(m)_{n-3}$, we can take $(m)_{n-3}$ to be max. Then we have

$$\left\langle \begin{matrix} [m']_{n-1} \\ (m)_{n-2} \end{matrix} \middle| \begin{matrix} W'_{n-1}, \\ \overset{\circ}{0} \end{matrix} \middle| \begin{matrix} [m]_{n-2} \\ (m)_{n-2} \end{matrix} \right\rangle = (W'_{n-1}!)^{1/2} \left[\frac{\mathcal{M}[m]_{n-2}}{\mathcal{M}[m']_{n-1}} \right]^{1/2},$$

$$(2.4)$$

$$\left\langle \begin{matrix} [m]_{n-1} \\ (m)_{n-2} \end{matrix} \middle| \begin{matrix} b, & \overset{\circ}{0} \\ & \overset{\circ}{0} \end{matrix} \middle| \begin{matrix} [m]_{n-2} \\ (m)_{n-2} \end{matrix} \right\rangle = (b!)^{1/2} \left[\frac{\mathcal{M}[m]_{n-2}}{\mathcal{M}[m]_{n-1}} \right]^{1/2}. \qquad (2.5)$$

Thus we have

$$\left\langle \begin{matrix} [m]_n \\ [m]_{n-1} \end{matrix} \middle| W_{n-1\,n} \middle| \begin{matrix} [m]_n \\ [m']_{n-1} \end{matrix} \right\rangle$$

$$= \left[\frac{\mathcal{M}[m]_n \mathcal{M}[m]_{n-2}}{\mathcal{M}[m']_{n-1} \mathcal{M}[m]_{n-1}} \right]^{1/2} \left\langle \begin{matrix} [m]_n \\ [m]_{n-1} \end{matrix} \middle\| \begin{matrix} p, & \overset{\circ}{0} \\ p, & \overset{\circ}{0} \end{matrix} \middle\| \begin{matrix} [m']_{n-1} \\ [m]_{n-2} \end{matrix} \right\rangle$$

$$= (-1)^{\epsilon_{n-1}[((m_{1\,n-1}-m_{n-1\,n-1})+(m'_{1\,n-1}-m'_{n-1\,n-1})+W'_n+W_n]}$$

$$\times [\dim[m']_{n-1} \dim[m]_{n-1}]^{1/2}$$

$$\times \left\{ \begin{matrix} [m_{1\,n}, \ldots, m_{n-1\,n}]\,[m_{1\,n-1}\cdots m_{n-1\,n-1}][W_n, \overset{\circ}{0}]_{n-1} \\ [m_{1\,n-2}\cdots m_{n-2\,n-2}\,0]\,[m'_{1\,n-1}\cdots m'_{n-1\,n-1}][\overset{\circ}{p}, 0]_{n-1} \end{matrix} \right\}.$$

$$(2.6)$$

The Weyl coefficient obviously occupies an important position in the finite transformation matrix of U(n). In the case of U(3) this has been calculated by many authors, among whom are Chacón and Moshinsky,[22] Holland,[23] Lezuo,[24] Majumdar and Basu,[25] Akyeampong and Rashid,[26] and Holman.[9] Our result agrees completely with Chacón and Moshinsky's, including the phase. Apart from phase, all other results agree with each other. Let us briefly mention the connection between these results.

Majumdar and Basu's 6-j symbol is the same as our (2.6) when we put $n=3$. Then we have

$$\left\langle \begin{matrix} m_{13} & m_{23} & 0 \\ m_{12} & m_{22} & \end{matrix} \middle\| W_{23} \middle\| \begin{matrix} m_{13} & m_{23} & 0 \\ m'_{12} & m'_{22} & \end{matrix} \right\rangle$$

$$= (-1)^{m_{12}+m'_{12}-m_{11}} [(m_{12}-m_{22}+1)(m'_{12}-m'_{12}-m'_{22}+1)]^{1/2}$$

$$\times \delta_{m_{13}+m_{23}-m_{12}-m_{22},\,m'_{12}+m'_{22}-m_{11}}$$

$$\times \left\{ \begin{matrix} \frac{1}{2}(m_{13}-m_{23}) & \frac{1}{2}(m_{12}-m_{22}) & \frac{1}{2}(m_{13}+m_{23}-m_{12}-m_{22}) \\ \frac{1}{2}m_{11} & \frac{1}{2}(m'_{12}-m'_{22}) & \frac{1}{2}(m_{12}+m_{22}-m_{11}) \end{matrix} \right\}.$$

$$(2.7)$$

Chacón and Moshinsky's expression is related to (2.7) through the Regge symmetry, and so is Akyeampong and Rashid's. Holland's result is equivalent to Akyeampong and Rashid's Eq. (2.5). Holman's expression is the same as (2.7), with the first row and second row interchanged (a difference in the definition of the 6-j symbol). Finally Lezuo's expression can be transformed into (2.7) in the following way. First write Lezuo's Eq. (13) as an

$$_4F_3 \left(\begin{matrix} b+f-d+1, & c+f-a+1, & f-a-c, f-b-d \\ f-e-a-d, e+f-a-d+1, & 2f+2 \end{matrix} \right) \quad (2.8)$$

with

$$a = \tfrac{1}{2}(\alpha_1+\alpha_2+\alpha_3+n) = \tfrac{1}{2}(m_{12}+m'_{12}-m_{23}-m_{11}),$$

$$b = \tfrac{1}{2}(\alpha_3+\alpha_{13}) = \tfrac{1}{2}(m_{13}+m_{23}-m_{12}-m_{22}),$$

$$c = \tfrac{1}{2}(\alpha_2+\alpha_{12}) = \tfrac{1}{2}(m_{12}+m_{22}-m_{11}),$$

$$d = \tfrac{1}{2}(\alpha_1+\alpha_{12}+\alpha_{13}+n) = \tfrac{1}{2}(m_{11}+m_{23}-m_{22}-m'_{22}), \quad (2.9)$$

$$e = \tfrac{1}{2}(\alpha_1+\alpha_2+\alpha_{13}-n) = \tfrac{1}{2}(m_{13}+m_{11}-m_{22}-m'_{12}),$$

$$f = \tfrac{1}{2}(\alpha_1+\alpha_3+\alpha_{12}+n) = \tfrac{1}{2}(m_{13}+m_{11}-m_{12}-m'_{22}).$$

Then (2.8) is related to $_4F_3(W,1)$ through Minton's[27] Eq. (7), and $_4F_3(W,1)$ is finally related to the 6-j symbol through Minton's Eq. (9). The 6-j symbol thus obtained is the same as Chacón and Moshinsky's. Thus all the results on the Weyl coefficients of U(3) are expressible as (2.7).

It is also clear that the Weyl coefficient is central to

the structure of the finite transformation matrix $D^{[m]}_{(m')\,n-1\,(m)\,n-1}$ in terms of Euler angles. From the parametrization of U(n) by Murnaghan[28] in terms of Euler angles and the work of Chacón and Moshinsky[22] on U(3), it is clear that all one has to know in obtaining the finite transformation matrix $D^{[m]}_{(m')\,n-1\,(m)\,n-1}$ is the Weyl coefficient $W_{n-1,n}$. The rest can be obtained by recurrence, or induction on n. In this regard it is interesting to observe that the finite transformation matrix for U(n) is a product of 6-j symbols in its subgroups and U(2) $D^j_{m'm}$ matrices.

In view of the significance of the labels z_i defined in the Appendix, which possess the following two properties:

(1) $z_i^* = -z_i,$ \hfill (2.10)

(2) $z_i^{(1)} + z_i^{(2)} = z_i^{(3)},$ \hfill (2.11)

it is desirable to construct the D matrices in terms of these labels. In the case of SU(3), these are just the isospin and hypercharge labels: $z_1 = 2I_z$, $z_2 = 3Y$. They have therefore direct application to physics. The D matrix diagonal in isospin and hypercharge in SU(3) has been obtained by Yeh.[11] It is

$$D(\alpha_1 \cdots \alpha_8) = \exp(iI_z\alpha_5)\exp(iY\alpha_7)\exp(iX_2\alpha_1)\exp(iI_z\alpha_4)$$

$$\times \exp(iX_7\alpha_2)\exp(iX_2\alpha_3)\exp(iI_z\alpha_6)\exp(iY\alpha_8),$$

$$(2.12)$$

where $X_1 \cdots X_8$ are the Gell-Mann[29] matrices. The finite transformation $D^{p,q}_{Imy|I'm'y'}(\alpha)$ is ($p=m_{13}-m_{23}, q=m_{23}$)

$$D^{p,q}_{Imy|I'm'y'}(\alpha)$$

$$= \exp[i(m\alpha_5 + m'\alpha_6 + y\alpha_7 + y'\alpha_8)] \sum_{m_1 I_1} N \exp(im_1\alpha_4)$$

$$\times d^I_{mm_1}(2\alpha_1) d^{I_1}_{-m_1/2+3y/4,\,-m_1/2+y'-y/4}(2\alpha_2)$$

$$\times d^{I'}_{m_1+y/2-y'/2,\,m'}(2\alpha_3), \quad (2.13)$$

where

$$N = (-1)^{2(p+q)/3+(y+y')/2-I-I'}(2I+1)^{1/2}(2I'+1)^{1/2}(2I_1+1)$$

$$\times W(a\,b\,c\,d;e\,f)W(a'\,b'\,c'\,d';e'\,f'),$$

$$a = (p+q)/6 - \tfrac{1}{2}y,$$

$$b = (2q-p)/6 - m_1/4 + y/8 + \tfrac{1}{2}(I+I_1),$$

$$c = (p-2q)/6 - y/8 + \tfrac{1}{2}(I+I_1) + m_1/4,$$

$$d = (p+q)/6 + y/4 + \tfrac{1}{2}m_1,$$

$$e = (2p-q)/6 - \tfrac{1}{2}I + y/8 + \tfrac{1}{2}I_1 - m_1/4,$$

$$f = (2p-q)/6 - \tfrac{1}{2}I_1 - m_1/4 + \tfrac{1}{2}I - y/4,$$

$$a' = (p+q)/6 + y/4 + \tfrac{1}{2}m_1,$$

$$b' = (2q-p)/6 - y/8 - m_1/4 + \tfrac{1}{2}(I_1+I') + y'/4,$$

$$c' = (p-2q)/6 + y/8 + \tfrac{1}{2}(I_1+I') + m_1/4 - y'/4,$$

$$d' = (p+q)/6 - \tfrac{1}{2}y',$$

$$e' = (2p-q)/6 - y/8 - \tfrac{1}{2}I_1 + \tfrac{1}{2}I' + \tfrac{1}{4}(y'-m_1),$$

$$f' = (2p-q)/6 - y/8 + \tfrac{1}{2}(I_1-I') + \tfrac{1}{4}(y'-m_1).$$

The D matrix can, of course, also be written as

1564   J. Math. Phys., Vol. 17, No. 8, August 1976

M.K.F. Wong   1564

boson polynomials. The relations between the bosons and the eight angles are as follows:

$$a_1^1 = [c_1 c_3 \exp(i\alpha_4/2) - s_1 c_2 s_3 \exp(-i\alpha_4/2)]$$
$$\times \exp\{i[\tfrac{1}{2}(\alpha_5 + \alpha_6) + (\alpha_7 + \alpha_8)/3]\},$$

$$a_1^3 = [c_1 s_3 \exp(i\alpha_4/2) + s_1 c_2 c_3 \exp(-\tfrac{1}{2}i\alpha_4)]$$
$$\times \exp\{i[\tfrac{1}{2}(\alpha_5 - \alpha_6) + (\alpha_7 + \alpha_8)/3]\},$$

$$a_1^2 = s_1 s_2 \exp\{i[\tfrac{1}{2}(-\alpha_4 + \alpha_5) + (\alpha_7 - 2\alpha_8)/3]\},$$

$$a_2^1 = -[s_1 c_3 \exp(i\tfrac{1}{2}\alpha_4) + c_1 c_2 s_3 \exp(-i\tfrac{1}{2}\alpha_4)]$$
$$\times \exp\{i[\tfrac{1}{2}(-\alpha_5 + \alpha_6) + (\alpha_7 + \alpha_8)/3]\},$$

$$a_2^2 = [-s_1 s_3 \exp(i\tfrac{1}{2}\alpha_4) + c_1 c_2 c_3 \exp(-i\tfrac{1}{2}\alpha_4)]$$
$$\times \exp\{i[\tfrac{1}{2}(-\alpha_5 - \alpha_6) + (\alpha_7 + \alpha_8)/3]\},$$

$$a_2^3 = c_1 s_2 \exp\{i[\tfrac{1}{2}(-\alpha_4 - \alpha_5) + (\alpha_7 - 2\alpha_8)/3]\},$$

$$a_3^1 = s_2 s_3 \exp\{i[\tfrac{1}{2}\alpha_6 + (-2\alpha_7 + \alpha_8)/3]\},$$

$$a_3^2 = -s_2 c_3 \exp\{i[-\tfrac{1}{2}\alpha_6 + (-2\alpha_7 + \alpha_8)/3]\},$$

$$a_3^3 = c_2 \exp[-i(2\alpha_7 + 2\alpha_8)/3], \text{ where } s_i = \sin\alpha_i, \ c_i = \cos\alpha_i.$$
$$\text{(2.14)}$$

The inverse relations are $(D_{ij} \equiv a_i^j)$

$$\cos^2 \alpha_1 = \frac{(D_{23}D_{31}D_{12} - D_{23}D_{32}D_{11})}{(1 - D_{33}D_{11}D_{22} + D_{33}D_{12}D_{21})},$$

$$\cos^2 \alpha_2 = D_{33}D_{11}D_{22} - D_{33}D_{12}D_{21},$$

$$\cos^2 \alpha_3 = (D_{32}D_{13}D_{21} - D_{32}D_{11}D_{23})/(1 - D_{33}D_{11}D_{22} + D_{33}D_{12}D_{21}),$$

$$\exp(i\alpha_4) = (iD_{13}D_{23})^{-1}D_{33}^{1/6}(D_{31}D_{32})^{-1/2}$$
$$\times (D_{23}D_{31}D_{12} - D_{23}D_{32}D_{11})^{1/2}$$
$$\times (D_{13}D_{32}D_{21} - D_{13}D_{31}D_{22})^{1/2}$$
$$\times (D_{12}D_{23} - D_{13}D_{22})^{1/2}(D_{11}D_{23} - D_{13}D_{21})^{1/2}$$
$$\times (D_{11}D_{22} - D_{12}D_{21})^{-1/2}$$
$$\times (1 - D_{33}D_{11}D_{22} + D_{33}D_{12}D_{21})^{-1/2},$$

$$\exp(i\alpha_5) = D_{13} \ D_{23} (D_{12}D_{23}D_{31} - D_{11}D_{23}D_{32})^{1/2}$$
$$\times (1 - D_{33}D_{11}D_{22} + D_{33}D_{12}D_{21})$$
$$\times (D_{13}D_{32}D_{21} - D_{13}D_{31}D_{22})^{-1/2},$$
$$\text{(2.15)}$$

$$\exp(i\alpha_6) = iD_{31}^{-1/2}D_{32}^{-3/2}D_{33}^{-2}(D_{12}D_{23} - D_{13}D_{22})^{1/2}$$
$$\times (D_{11}D_{23} - D_{13}D_{21})^{3/2}$$

$$\exp(i\alpha_7) = D_{33}^{-3/4}i^{-1/2}D_{31}^{-1/4}D_{32}^{-1/4}(D_{11}D_{22} - D_{12}D_{21})^{1/4}$$
$$\times (D_{12}D_{23} - D_{13}D_{22})^{1/4}(D_{11}D_{23} - D_{13}D_{21})^{1/4},$$

$$\exp(i\alpha_8) = i^{1/2}D_{33}^{-1/2}D_{31}^{1/4}D_{32}^{1/4}(D_{11}D_{22} - D_{12}D_{21})^{1/2}$$
$$\times (D_{12}D_{23} - D_{13}D_{22})^{-1/4} \ (D_{11}D_{23} - D_{13}D_{21})^{-1/4}.$$

Thus the structure of the boson polynomial in $U(n) * U(n)$ is further clarified. Let us make one more remark on $U(2)$. That is, the Wigner coefficient of $U(2)$ is itself contained in the boson polynomial of $U(2) * U(2)$.

Proof: The normalized boson polynomial in $U(2) * U(2)$ is

$$\mathcal{N}^{-1/2} B \begin{pmatrix} \mu_{11} \\ m_{12} m_{12} \\ m_{11} \end{pmatrix} = \sum_s \frac{[(m_{11} - m_{22})! (m_{12} - m_{11})! (\mu_{11} - m_{22})! (m_{12} - \mu_{11})!]^{1/2}}{s! (\mu_{11} - m_{22} - s)! (m_{11} - m_{22} - s)! (m_{12} - m_{11} - \mu_{11} + m_{22} + s)!}$$

$$\times \left[\frac{(m_{12} + 1)! m_{22}!}{(m_{12} - m_{22} + 1)} \right]^{-1/2} (a_1^1 a_2^2 - a_1^2 a_2^1)^{m_{22}} a_1^{1 s} a_1^{2\mu_{11} - m_{22} - s} a_2^{1 m_{11} - m_{22} - s} a_2^{2 m_{12} - m_{11} - \mu_{11} + m_{22} + s}$$

$$= \sum_{s, x} \left\{ \frac{[(m_{11} - m_{22})! (m_{12} - m_{11})! (\mu_{11} - m_{22})! (m_{12} - \mu_{11})! m_{22}! (m_{12} - m_{22} + 1)]^{1/2}(-1)^x}{s! (\mu_{11} - m_{22} - s)! (m_{11} - m_{22} - s)! (m_{12} - m_{11} - \mu_{11} + m_{22} + s)! x! (m_{22} - x)!} \right.$$

$$\times [(m_{12} + 1)!]^{-1/2}[(x + s)! (m_{11} - x - s)! (\mu_{11} - x - s)! (m_{12} + m_{22} - m_{11} - \mu_{11} + x + s)!]^{1/2} \Big\}$$

$$\times \frac{a_1^{1 x+s} a_1^{2 m_{11} - (x+s)} a_2^{1 \mu_{11} - (x+s)} a_2^{2 m_{12} + m_{22} - m_{11} - \mu_{11} + x + s}}{[(x + s)! (m_{11} - x - s)! (\mu_{11} - x - s)! (m_{12} + m_{22} - m_{11} - \mu_{11} + x + s)!]^{1/2}} . \tag{2.16}$$

Now putting $\alpha = x + s$, we find that $\sum_x \{\ \}$ in Eq. (2.16) is the Wigner coefficient in $U(2)$,

$$c_{m_1 m_2 m_3}^{j_1 j_2 j_3} = c_{m_{11} - \alpha - (m_{12} + m_{22} - \mu_{11})/2, \ \alpha - \mu_{11}/2, \ m_{11} - (m_{12} + m_{22})/2}^{(m_{11} + m_{22} - \mu_{11})/2, \ \mu_{11}/2, \ (m_{12} - m_{22})/2} = \left\langle \begin{array}{cc} m_{12} \ m_{22} \\ m_{11} \end{array} \Big| \begin{array}{cc} \mu_{11} \ 0 \\ \alpha \end{array} \Big| \begin{array}{cc} m_{12} + m_{22} - \mu_{11}, \ 0 \\ m_{11} - \alpha \end{array} \right\rangle$$

$$= (-1)^{m_{22}} \left\langle \begin{array}{cc} m_{12} \ m_{22} \\ m_{11} \end{array} \Big| \begin{array}{cc} W_2', \ 0 \\ m_{11} - \alpha \end{array} \Big| \begin{array}{cc} \mu_{11} \ 0 \\ \alpha \end{array} \right\rangle.$$

Thus knowing the double boson polynomial $U(2) * U(2)$, one can write down the Wigner coefficient of $U(2)$ by inspection. This is certainly the easiest way of deriving the Wigner coefficients of $U(2)$. If one wishes to write the expression $c_{m_1 m_2 m_1 + m_2}^{j_1 j_2 j_3}$ one merely has to make the following identification:

$$m_{12} = j_1 + j_2 + j_3, \quad m_{22} = j_1 + j_2 - j_3, \quad m_{11} = j_1 + j_2 + m_1 + m_2, \quad \mu_{11} = 2j_2, \quad \alpha = j_2 + m_2 \tag{2.17}$$

One then finds that $c_{m_1 m_2 m_1 + m_2}^{j_1 j_2 j_3}$ as given by $\sum_x \{\ \}$ in Eq. (2.16) is identical to the expression given by Edmonds,[14] Eq. (3.6.11).

3. STRETCHED 6-j SYMBOLS OF U(n)

In this section we shall derive an expression for the stretched 6-j symbols of U(n). This is done by putting $q = 0$ in Eq. (1.17). We then have

$$\left\langle \begin{matrix} [m_{1n}\circ\circ\circ m_{n-1\,n}, & 0] \\ [m_{1\,n-1}\circ\circ\circ m_{n-1\,n-1}] \end{matrix} \left\| \begin{matrix} p, & \dot{0} \\ & 0 \end{matrix} \right\| \begin{matrix} [m'_{1n}\circ\circ\circ m'_{n-1\,n}, & 0] \\ [m_{1\,n+1}\cdots m_{n-1\,n-1}] \end{matrix} \right\rangle = (-1)^{y([m]_{n-1})+y([m]_{n'}} \left[\frac{\mathcal{M}[m']_n W_n!}{\mathcal{M}[m]_n W'_n!} \right]^{1/2} [\dim[m]_{n-1} \dim[W_n, \dot{0}]_{n-1}]^{1/2}$$

$$\times (-1)^{\epsilon_{n-1}(p+m_{1\,n-1}-m_{n-1\,n-1}+m_{1\,n}-m_{n-1\,n-1}+W'_n)} \left\{ \begin{matrix} [m_{1\,n-1}\cdots m_{n-1\,n-1}] & [m'_{1n}\cdots m'_{n-1\,n}] & [W'_n, \dot{0}]_{n-1} \\ [\dot{p}, 0]_{n-1} & [W_n, \dot{0}]_{n-1} & [m_{1n}\cdots m_{n-1\,n}] \end{matrix} \right\}. \tag{3.1}$$

The 6-j symbol in (3.1) is stretched because $p + W'_n = W_n$.

Now the left-hand side of Eq. (3.1) has been evaluated by Chacón et al.[7] and the result is

$$(p!)^{1/2} \frac{S_{n\,n-1}(m_{1n}\cdots m_{n-1\,n}, 0; m_{1\,n-1}\circ\circ\circ m_{n-1\,n-1}) S_{nn}(m_{1n}\cdots m_{n-1\,n}, 0; m_{1n}\cdots m_{n-1\,n}, 0)}{S_{n\,n-1}(m'_{1n}\cdots m'_{n-1\,n}, 0; m_{1\,n-1}\cdots m_{n-1\,n-1}) S_{nn}(m_{1n}\cdots m_{n-1\,n}, 0; m'_{1n}\cdots m'_{n-1\,n}, 0)}, \tag{3.2}$$

where

$$S_{nm}(h_1\cdots h_n; q_1\cdots q_m) \equiv \left[\prod_{k=1}^{m}\prod_{s=1}^{k}(h_s - q_k + k - s)! \Big/ \prod_{k=1}^{n-1}\prod_{s=k+1}^{n}(q_k - h_s + s - k - 1)! \right]^{1/2}. \tag{3.3}$$

Equating (3.1) and (3.2), we obtain the stretched 6-j symbol in U($n-1$):

$$\left\{ \begin{matrix} [m_{1\,n-1}\cdots m_{n-1\,n-1}] & [m'_{1n}\circ\circ\circ m'_{n-1\,n}] & [W'_n, 0]_{n-1} \\ [\dot{p}, 0]_{n-1} & [W_n, 0]_{n-1} & [m_{1n}\cdots\circ m_{n-1\,n}] \end{matrix} \right\} = (-1)^{y([m]_{n-1})+y([m]_n)}(-1)^{\epsilon_{n-1}(\Sigma_{i=2}^{n-2}m_{in}+\Sigma_{i=2}^{n-2}m_{i\,n-1}+2m_{1n}-2m_{n-1\,n-1})}$$

$$\times \left(\frac{p! W'_n!}{W_n!} \right)^{1/2} (\dim[m'_{1n}\cdots m'_{n-1\,n}] \dim[W_n, \dot{0}]_{n-1})^{-1/2} \frac{S_{n\,n-1}(m'_{1n}\cdots m'_{n-1\,n}, 0; m'_{1n}\circ\circ\circ m'_{n-1\,n})}{S_{n\,n-1}(m'_{1n}\cdots m'_{n-1\,n}, 0; m_{1\,n-1}\cdots m_{n-1\,n-1})}$$

$$\times \frac{S_{n\,n-1}(m_{1n}\cdots m_{n-1\,n}, 0; m_{1\,n-1\,n-1})\circ\circ\circ m_{n-1\,n-1})}{S_{n\,n-1}(m_{1n}\cdots m_{n-1\,n}, 0; m'_{1n}\cdots m'_{n-1\,n})}. \tag{3.4}$$

In the case of U(2), ($n=3$), Eq. (3.4) gives

$$\left\{ \begin{matrix} \frac{1}{2}(m_{12}-m_{22}) & \frac{1}{2}(m'_{13}-m'_{23}) & \frac{1}{2}(m'_{13}+m'_{23}-m_{12}-m_{22}) \\ \frac{1}{2}(m_{13}+m_{23}-m'_{13}-m'_{23}) & \frac{1}{2}(m'_{13}+m'_{23}-m_{12}-m_{22}) & \frac{1}{2}(m_{13}-m_{23}) \end{matrix} \right\}$$

$$= \left\{ \begin{matrix} \frac{1}{2}(m_{12}-m_{22}) & \frac{1}{2}(m_{13}-m_{23}) & \frac{1}{2}(m_{13}+m_{23}-m_{12}-m_{22}) \\ \frac{1}{2}(m_{13}+m_{23}-m'_{13}-m'_{23}) & \frac{1}{2}(m'_{13}+m'_{23}-m_{12}-m_{22}) & \frac{1}{2}(m'_{13}-m'_{23}) \end{matrix} \right\} \equiv \left\{ \begin{matrix} j_1 & j_2 & l_1+l_2 \\ l_1 & l_2 & l_3 \end{matrix} \right\}$$

$$= (-1)^{m_{13}+m_{22}} \left[\frac{(2l_1)!(2l_2)!}{(2l_1+2l_2+1)!(2l_3+1)} \right]^{1/2} \frac{S_{32}(m'_{13}m'_{23}0; m'_{13}m'_{23})S_{32}(m_{13}m_{23}0; m_{12}m_{22})}{S_{32}(m_{13}m_{23}0; m'_{13}m'_{23})S_{32}(m'_{13}m'_{23}0; m_{12}m_{22})}$$

$$= (-1)^{m_{13}+m_{22}} \left[\frac{(2l_1)!(2l_2)!(m_{13}-m_{12})!(m_{13}-m_{22}+1)!(m_{23}-m_{22})!(m'_{13}-m_{23})!(m_{12}-m'_{23})!}{(2l_1+2l_2+1)!(m_{12}-m_{23})!(m_{13}-m'_{13})!(m_{13}-m'_{23}+1)!(m_{23}-m'_{23})!(m'_{13}-m_{12})!} \right]^{1/2}$$

$$\times \left[\frac{1}{(m'_{13}-m_{22}+1)!(m'_{23}-m_{22})!} \right]^{1/2}. \tag{3.5}$$

Equation (3.5) agrees with Eq. (6.3.1) of Edmonds.[14]

Next let us compare the stretched 6-j symbol of U(n) with the generalized beta function of Gel'fand and Graev.[10] It has been shown by Louck and Biedenharn[16c] that the generalized beta functions can be written in terms of isoscalar factors:

$$B\begin{pmatrix} (m') \\ [m] \\ (m) \end{pmatrix}(I + te_{n\,n-1}) = B\begin{pmatrix} (m) \\ [m] \\ (m') \end{pmatrix}(I + te_{n-1\,n})$$

$$= \delta_{(m)_{n-2}(m')_{n-2}} \left[\frac{\mathcal{M}([m]_n)}{b!\,b'!\mathcal{M}([m]_{n-1})} \right]^{1/2} \left\langle \begin{matrix} [m]_n & [b, \dot{0}]_n & [m']_{n-1}, 0 \\ [m]_{n-1} & 0 & [m]_{n-1} \end{matrix} \right\rangle \left\langle \begin{matrix} [m']_{n-1} & [b', \dot{0}]_{n-1} & [m]_{n-1} \\ [m]_{n-2} & 0 & [m]_{n-1} \end{matrix} \right\rangle t^{b'}. \tag{3.6}$$

Without loss of generality, we can put $m_{nn} = 0$, then

$$\left\langle \begin{array}{c} [m_{1\,n} \cdots m_{n-1\,n}, 0] \| [b, \dot{0}]_n \| [m']_{n-1}, 0 \\ [m]_{n-1} \quad \| \quad 0 \quad \| \quad [m]_{n-1} \end{array} \right\rangle$$

$$= (-1)^{y([m]_{n-1}) + y([m]_n)} \left[\frac{\mathcal{M}[m']_{n-1} W_n!}{\mathcal{M}[m]_n W_n'!} \right]^{1/2} (-1)^{\epsilon_{n-1} (\Sigma_{i=2}^{n-2} m_{in} + \Sigma_{i=2}^{n-2} m_{i\,n-1} + 2m_{1n} - 2m_{n-1\,n-1})} (\dim[m]_{n-1} \dim[W_n, \dot{0}]_{n-1})^{1/2}$$

$$\times \left\langle \begin{array}{ccc} [m_{1\,n-1} \cdots m_{n-1\,n-1}] & [m'_{1\,n-1} \cdots m'_{n-1\,n-1}] & [W'_n, \dot{0}]_{n-1} \\ [\dot{b}, 0]_{n-1} & [W_n, \dot{0}]_{n-1} & [m_{1\,n} \cdots m_{n-1\,n}] \end{array} \right\rangle . \tag{3.7}$$

The other term is an isoscalar factor in $U(n-1)$, whose value is equal to

$$\left\langle \begin{array}{c} [m']_{n-1} \| [b', \dot{0}]_{n-1} \| [m]_{n-1} \\ [m]_{n-2} \quad \| \quad 0 \quad \| \quad [m]_{n-2} \end{array} \right\rangle$$

$$= (b'!)^{1/2} \frac{S_{n-1\,n-2}(m'_{1\,n-1} \cdots m'_{n-1\,n-1}; m_{1\,n-2} \cdots m_{n-2\,n-2})}{S_{n-1\,n-2}(m_{1\,n-1} \cdots m_{n-1\,n-1}; m_{1\,n-2} \cdots m_{n-2\,n-2})} \frac{S_{n-1\,n-1}(m'_{1\,n-1} \cdots m'_{n-1\,n-1}; m'_{1\,n-1} \cdots m'_{n-1\,n-1})}{S_{n-1\,n-1}(m'_{1\,n-1} \cdots m'_{n-1\,n-1}; m_{1\,n-1} \cdots m_{n-1\,n-1})} . \tag{3.8}$$

Thus the generalized beta function is connected to the stretched 6-j symbol of $U(n-1)$ through Eqs. (3.6), (3.7), and (3.8). Many of the terms in Eq. (3.8) are similar to the terms in the stretched 6-j symbols of $U(n-2)$. However, they cannot be entirely identified; so we shall leave (3.8) as it is. Thus we conclude that the generalized beta functions of Gel'fand and Graev can be written either as the product of two isoscalar factors, one in $U(n)$ and one in $U(n-1)$, or as the product of a stretched 6-j symbol in $U(n-1)$ and an isoscalar factor in $U(n-1)$.

4. REGGE SYMMETRY OF 6-j SYMBOLS OF U(2) AND WEYL COEFFICIENTS OF U(3)

Louck and Biedenharn[16c] have shown that the Regge symmetry of 3-j symbols of $U(2)$ are connected with the boson polynomials of both $U(2)$ and $U(3)$. For $U(2)$, the relation is

$$C \left(\begin{array}{ccc} j_1 + j_2 + j & j_1 + j_2 + m & j_1 + j_2 - j \\ & 2j_1 & \end{array} \right) \left(\begin{array}{cc} j_1 + m_1 & j_1 - m_1 \\ j_2 + m_2 & j_2 - m_2 \end{array} \right)$$

$$= C^{j_1 j_2 j_3}_{m_1 m_2 m} \tag{4.1}$$

Then the transpositional symmetry $j_1 - m_1 \leftrightarrow j_2 + m_2$ or $m_1 + m_2 \leftrightarrow j_1 - j_2$ is the Regge symmetry as discussed by Bincer.[19]

For $U(3)$, the relation is

$$C \left(\begin{array}{ccccc} & & k & & \\ & k & & k & \\ k & & k & & k \\ & k & & k & \\ & & k & & \end{array} \right) \left(\begin{array}{ccc} j_1 + m_1 & j_1 - m_1 & -j_1 + j_2 + j \\ j_2 + m_2 & j_2 - m_2 & j_1 - j_2 + j \\ j - m & j + m & j_1 + j_2 - j \end{array} \right)$$

$$= (-1)^{j_1 - j_2 + m} \left[\frac{2(k!)}{(2j+1)(k+2)!} \right]^{1/2} C^{j_1 j_2 j}_{m_1 m_2 m} \tag{4.2}$$

with $j_1 + j_2 + j = k$.

A similar question can be asked about the Regge symmetry of the 6-j symbols of $U(2)$. As Jahn and Howell[30] pointed out, the 144 symmetries of the 6-j symbol can best be understood if one writes the 6-j symbol $\left\{ \begin{array}{ccc} j_1 & j_2 & j_3 \\ k_1 & k_2 & k_3 \end{array} \right\}$ as

$$\left[\begin{array}{cccc} J_0 & J_1 & J_2 & J_3 \\ & K_1 & K_2 & K_3 \end{array} \right]$$

$$\equiv \left\{ \begin{array}{ccc} \frac{1}{2}(J_0 + J_1 - K_1) & \frac{1}{2}(J_0 + J_2 - K_2) & \frac{1}{2}(J_0 + J_3 - K_3) \\ \frac{1}{2}(J_2 + J_3 - K_1) & \frac{1}{2}(J_1 + J_3 - K_2) & \frac{1}{2}(J_1 + J_2 - K_3) \end{array} \right\} \tag{4.3}$$

with

$$J_0 + J_1 + J_2 + J_3 - K_1 - K_2 - K_3 = 0,$$

$$J_0 = j_1 + j_2 + j_3, \quad J_1 = j_1 + k_2 + k_3, \quad J_2 = j_2 + k_1 + k_3,$$

$$J_3 = j_3 + k_1 + k_2, \quad K_1 = j_2 + j_3 + k_2 + k_3,$$

$$K_2 = j_1 + j_3 + k_1 + k_3, \quad K_3 = j_1 + j_2 + k_1 + k_2. \tag{4.4}$$

Or, conversely,

$$j_1 = \tfrac{1}{2}(J_0 + J_1 - K_1), \quad j_2 = \tfrac{1}{2}(J_0 + J_2 - K_2), \quad j_3 = \tfrac{1}{2}(J_0 + J_3 - K_3),$$

$$k_1 + \tfrac{1}{2}(J_2 + J_3 - K_1), \quad k_2 = \tfrac{1}{2}(J_1 + J_3 - K_2), \quad k_3 = \tfrac{1}{2}(J_1 + J_2 - K_3), \tag{4.5}$$

the Regge symmetry is then composed of the products of separate permutations of J_0, J_1, J_2, and J_3 with separate permutations of K_1, K_2, and K_3.

Since the 6-j symbol of $U(2)$ is connected with the Weyl coefficient of $U(3)$, we now investigate the symmetry of the Weyl coefficient of $U(3)$. We have seen that the Weyl coefficient of $U(3)$, W_{23}, imposes the constraint

$$m_{11} = m'_{11}, \quad m_{13} + m_{23} - m_{12} - m_{22} - m'_{12} - m'_{22} + m_{11} = 0. \tag{4.6}$$

We can now make the following identification:

$$W_{23} \text{ of } \left| \begin{array}{ccc} m_{11} & & \\ m_{12} & m'_{22} & \\ m_{13} & m_{23} & 0 \\ m_{12} & m_{22} & \\ m_{11} & & \end{array} \right\rangle = W_{23} \text{ of } \left| \begin{array}{ccc} J_2 + J_3 - K_1 & & \\ J_2 & K_2 - J_0 & \\ K_2 & J_2 + J_3 - K_3 & 0 \\ J_3 & K_2 - J_1 & \\ J_2 + J_3 - K_1 & & \end{array} \right\rangle$$

1567   J. Math. Phys., Vol. 17, No. 8, August 1976

M.K.F. Wong   1567

$$= (-1)^{m_{12}+m'_{12}-m_{11}}[(m_{12}-m_{22}+1)(m'_{12}-m'_{22}+1)]^{1/2}$$

$$\times \left\{ \begin{array}{ccc} \frac{1}{2}(m_{13}-m_{23}) & \frac{1}{2}(m_{12}-m_{22}) & \frac{1}{2}(m_{13}+m_{23}-m_{12}-m_{22}) \\ \frac{1}{2}m_{11} & \frac{1}{2}(m'_{12}-m'_{22}) & \frac{1}{2}(m_{12}+m_{22}-m_{11}) \end{array} \right\}$$

$$= (-1)^{K_1}(J_3-K_2+K_1+1)^{1/2}(J_2-K_2+J_0)^{1/2} \left[\begin{array}{cccc} J_0 & J_1 & J_2 & J_3 \\ & K_1 & K_2 & K_3 \end{array} \right]$$

$$= (-1)^{j_2+j_3+k_2+k_3}[(2k_2+1)(2j_2+1)]^{1/2} \left\{ \begin{array}{ccc} j_1 & j_2 & j_3 \\ k_1 & k_2 & k_3 \end{array} \right\}, \qquad (4.7)$$

where W_{23} is the boson

$$\begin{pmatrix} 1 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \end{pmatrix}.$$

It is clear that, as far as the Weyl coefficient is concerned, the labels of the J and K in the state

$$\left| \begin{array}{c} J_2+J_3-K_1 \\ J_2 K_2 - J_0 \\ K_2 J_2 + J_3 - K_3 \; 0 \\ J_3 K_2 - J_1 \\ J_2 + J_3 - K_1 \end{array} \right\rangle$$

are separately permutable. Thus we obtain the $3! \times 4! = 144$ symmetries of Regge from this state. Moreover, the constraint equation (4.6) is just equal to the constraint imposed by Jahn and Howell:

$$J_0 + J_1 + J_2 + J_3 - K_1 - K_2 - K_3 = 0. \qquad (4.8)$$

ACKNOWLEDGMENT

APPENDIX: PHASE RELATIONS BETWEEN WIGNER COEFFICIENTS AND 3-j SYMBOLS OF U(n)

The phase relations of Wigner coefficients are closely connected with R conjugation. The R conjugation has been discussed very thoroughly by Baird and Biedenharn.[13] We shall denote the conjugate representation by star *. Thus, for U(n)

$$m_{ij}^* = m_{1n} - m_{j-i+1,j}. \qquad (A1)$$

All relations between Wigner coefficients themselves as well as between Wigner coefficients and 3-j symbols can be derived from the two relations:

$$(m^{(1)},m^{(2)},m^{(3)}) = (-1)^r \{m^{(1)},m^{(2)},m^{(3)*}\}, \qquad (A2)$$

$$\{m^{(1)},m^{(2)},m^{(3)}\} = (-1)^s \{m^{(1)*},m^{(2)*},m^{(3)*}\}$$

$$= (-1)^s \text{ odd permutation of columns of } \{m^{(1)},m^{(2)}m^{(3)}\}. \qquad (A3)$$

In the case of U(2),

$$r = j_1 - j_2 + m_1 + m_2, \qquad (A4)$$

$$s = j_1 + j_2 + j_3. \qquad (A5)$$

Clearly what one has to do in U(n) is to extend the labels j and m to higher groups, so that they retain the same properties under R conjugation. These labels are not arbitrary at all. In U(2), under R conjugation, we have $j^* = j$, $m^* = -m$. In U(3), the corresponding labels are $\lambda = m_{13} - m_{23}$, $\mu = m_{23}$, $z_1 \equiv 2m = 2m_{11} - (m_{12}+m_{22})$, $z_2 \equiv 3Y = 3(m_{12}+m_{22}) - 2(m_{13}+m_{23})$, so that, under conjugation,

$$\lambda^* = \mu, \quad \mu^* = \lambda, \quad z_1^* = -z_1, \quad z_2^* = -z_2. \qquad (A6)$$

Thus we can define the following labels:

$$z_1 \equiv 2m = 2m_{11} - (m_{12}+m_{22}),$$

$$z_2 \equiv 3Y = 3(m_{12}+m_{22}) - 2(m_{13}+m_{23}+m_{33}),$$

$$z_3 = 4(m_{13}+m_{23}+m_{33}) - 3(m_{14}+m_{24}+m_{34}+m_{44}),$$

...

$$z_i = (i+1)\sum_{j=1}^{i} m_{ji} - i\sum_{j=1}^{i+1} m_{j,i+1}. \qquad (A7)$$

Under R conjugation, we have

$$z_i^* = -z_i \qquad (A8)$$

Moreover, since the z_i's are diagonal operators, they have the property

$$z_i^{(1)} + z_i^{(2)} = z_i^{(3)}. \qquad (A9)$$

Next we define the following labels, corresponding to $\lambda(=m_{13}-m_{23})$ and $\mu(=m_{23})$ in SU(3):

$$\lambda_1 = m_{1n} - m_{2n}, \quad \lambda_2 = m_{2n} - m_{3n}, \ldots, \lambda_i = m_{in} - m_{i+1,n})$$

$$\lambda_n = 0 \qquad (A10)$$

These labels have the following property under R conjugation:

$$\lambda_1^* = \lambda_{n-1}, \quad \lambda_2^* = \lambda_{n-2}, \ldots, \lambda_{n-1}^* = \lambda_1, \quad \lambda_n^* = \lambda_n = 0. \qquad (A11)$$

We can now extend the phase relations in (A2) and (A3) to U(n):

$$(m^{(1)},m^{(2)},m^{(3)})$$

$$= (-1)^{(z_1^{(1)}+z_2^{(1)}+\cdots+z_{n-1}^{(1)})/2+(z_1^{(2)}+z_2^{(2)}+\cdots+z_{n-1}^{(2)})/2}$$

$$(-1)^{\epsilon_n(m_{1n}^{(1)}-m_{nn}^{(1)})-\epsilon_n(m_{1n}^{(2)}-m_{nn}^{(2)})}\{m^{(1)},m^{(2)},m^{(3)*}\}, \qquad (A12)$$

$$\{m^{(1)},m^{(2)},m^{(3)}\} = (-1)^{\epsilon_n(m_{1n}^{(1)}-m_{nn}^{(1)}+m_{1n}^{(2)}-m_{nn}^{(2)}+m_{1n}^{(3)}-m_{nn}^{(3)})}$$

$$\{m^{(1)*},m^{(2)*},m^{(3)*}\} = (-1)^{\epsilon_n(m_{1n}^{(1)}-m_{nn}^{(1)}+m_{1n}^{(2)}-m_{nn}^{(2)}+m_{1n}^{(3)}-m_{nn}^{(3)})}$$

odd permutation of columns of $\{m^{(1)},m^{(2)},m^{(3)}\}$.

$$\qquad (A13)$$

where

$$\epsilon_n = \tfrac{1}{2} \text{ for } n = 2 + 4k$$

$$1 \text{ for } n = 3 + 4k$$

$$\tfrac{1}{2} \text{ for } n = 4 + 4k$$

$$0 \text{ for } n = 5 + 4k, \quad k = 0,1,2,\cdots. \qquad (A14)$$

In the case of U(3), we obtain

$$(m^{(1)}, m^{(2)}, m^{(3)}) = (-1)^{m_{11}^{(1)} + m_{12}^{(1)} + m_{22}^{(1)} - m_{23}^{(1)}}$$

$$\times \left[\frac{\dim m^{(3)}}{\dim m^{(2)}}\right]^{1/2} (m^{(1)}, m^{(3)*}, m^{(2)*}) \qquad (A15)$$

This phase agrees with Baird and Biedenharn for R conjugation, i.e., $(-1)^{\rho(m)}{}_{n-1}{}^{-\rho(m)}{}_{\max}$ in their notation. It should be noted that this phase differs from de Swart[12] (or Resnikoff[5]) and Ponzano.[4] However, we can use Resnikoff's method to obtain our phase if we define the basis state as

$$|\lambda\mu, \alpha\rangle = N(\lambda\mu;\alpha) \sum_k \binom{r}{k} \frac{(\mu - q)! \, p!}{(\mu - q - k)!(p - r + k)!} \xi_1^{p-(r-k)} \eta_1^{r-k}$$

$$\sigma_1^{\lambda-p} (\delta_{12}^{(1)})^k (-\delta_{12}^{(2)})^{\mu-q-k} (\delta_{12}^{(3)})^q \qquad (A16)$$

Since (A16) is the more natural definition of the basis state and is, in fact, the one used by many authors, e.g., Akyeampong and Rashid,[26] Majumdar and Basu,[25] we suggest that the phase of the "1-j" symbol in SU(3) should be written as in (A15).

Baird and Biedenharn[13] stated that an over-all phase is arbitrary in U(n) R conjugation. In view of (A12) and (A13), we think that even that over-all phase can be determined for U(n), and is therefore not completely arbitrary.

Applying Eqs. (A12) and (A13) to Eqs. (1.3), (1.4), and (1.6) in the text, we obtain, e.g.,

$$\left((k)_n, \binom{c, \ 0}{0}_n, (d)_n\right) \equiv (m^{(2)}, m^{(1)}, m^{(3)})$$

$$= (-1)^y \left(\binom{c, \ 0}{c}_n, (d)_n, (k)_n\right),$$

where

$$y = -\frac{1}{2} \sum_{i=1}^{n-1} z_i^{(1)} + \sum_{i=1}^{n-1} z_i^{(3)} + 3\epsilon_n(m_{1n}^{(1)} - m_{nn}^{(1)}) + 2\epsilon_n(m_{1n}^{(3)} - m_{nn}^{(3)})$$

with $\qquad\qquad\qquad\qquad\qquad\qquad\qquad$ (A17)

$$z_i^{(1)} + z_i^{(2)} = z_i^{(3)},$$

$y = 0$ for U(2), U(3), U(5), ..., U($2k+1$),

$$\sum_{i=2}^{n-1} m_{in} \quad \text{for U(4), U(6), \ldots, U(2k).}$$

Thus y is a function, if nonvanishing, of $[d]_n$ only, where

$$[d]_n \equiv m^{(3)} = [m_{1n}, \ldots, m_{nn}]. \qquad (A18)$$

The other cases are similar to the one discussed above.

[1]M. Moshinsky, Rev. Mod. Phys. 34, 813 (1962).
[2]L. Banyai, N. Marinescu, I. Raszillier, and V. Rittenberg, Phys. Lett. 14, 156 (1965).
[3]Hou Pei-Yu, Scientia Sinica 14, 367 (1965).
[4]G. Ponzano, Nuovo Cimento A 41, 142 (1966).
[5]M. Resnikoff, J. Math. Phys. 8, 63 (1967).
[6]S. Alisauskas, Liet. Fiz. Rink. 9, 641 (1969).
[7]E. Chacón, M. Ciftan, and L.C. Biedenharn, J. Math. Phys. 13, 577 (1972).
[8]S.J. Alisauskas, A.A.A. Jucys, and A.P. Jucys, J. Math. Phys. 13, 1329 (1972).
[9]W.J. Holman III, Nuovo Cimento A 4, 904 (1971).
[10]I.M. Gel'fand and M.I. Graev, Izv. Akad. Nauk SSSR Mekh. Mashinostr. 29, 1329 (1965) [Am. Math. Soc. Transl. Ser. 2 65, 116 (1967)].
[11]H.Y. Yeh, Ann. N.Y. Acad. Sci. 172, 721 (1971).
[12]J.J. de Swart, Rev. Mod. Phys. 35, 916 (1963).
[13]G.E. Baird and L.C. Biedenharn, J. Math. Phys. 5, 1723 (1964).
[14]A.R. Edmonds, Angular Momentum in Quantum Mechanics (Princeton U.P., Princeton, N.J., 1960).
[15]R.T. Sharp and H. von Baeyer, J. Math. Phys. 7, 1105 (1966).
[16](a) L.C. Biedenharn, J.D. Louck, E. Chacón, and M. Ciftan, J. Math. Phys. 13, 1957 (1972); (b) L.C. Biedenharn, and J.D. Louck, J. Math. Phys. 13, 1985 (1972); (c) J.D. Louck and L.C. Biedenharn, J. Math. Phys. 14, 1336 (1973).
[17]L.C. Biedenharn, A. Giovannini, and J.D. Louck, J. Math. Phys. 8, 691 (1967).
[18]A.A. Jucys, Liet. Fiz. Rink. 10, 5 (1970).
[19]A. Bincer, J. Math. Phys. 11, 1835 (1970).
[20]T. Regge, Nuovo Cimento 10, 544 (1958).
[21]T. Regge, Nuovo Cimento 11, 116 (1959).
[22]E. Chacón and M. Moshinsky, Phys. Lett. 23, 567 (1966).
[23]D.F. Holland, J. Math. Phys. 10, 531 (1969).
[24]J. Lezuo, J. Math. Phys. 8, 1163 (1967).
[25]S.D. Majumdar and B.K. Basu, Ann. Phys. (N.Y.) 56, 464 (1970).
[26]D.A. Akyeampong and M.A. Rashid, J. Math. Phys. 13, 1218 (1972).
[27]B.M. Minton, J. Math. Phys. 11, 3061 (1970).
[28]F.D. Murnaghan, The Unitary and Rotation Groups (Spartan, Washington, D.C., 1962).
[29]M.A. Gell-Mann and Y. Néeman, The Eightfold Way (Benjamin, New York, 1964).
[30]H.A. Jahn and K.M. Howell, Proc. Camb. Phil. Soc. 55, 338 (1959).

A "nonstandard" approach to the thermodynamic limit. II. Weakly tempered potentials and neutral Coulomb systems*

A. Ostebee, P. Gambardella, and M. Dresden

Institute for Theoretical Physics, State University of New York at Stony Brook, Stony Brook, New York 11794
(Received 22 January 1976)

Nonstandard analysis is used to prove the existence and uniqueness of the thermodynamic limit for a number of physical systems. Included are neutral Coulomb systems and systems described by weakly tempered potentials. The basic feature of the method is the application of the transfer theorem of nonstandard analysis to the inequalities for the free energy for finite systems. Nonstandard transcriptions of packing theorems are also needed. The nonstandard procedure has the great advantage of providing an explicit separation of the physical and geometric elements in the proof. This allows a unified treatment of several distinct cases, so that the demonstrations—in spite of unavoidable technical details—become physically intuitive and mathematically straightforward.

I. INTRODUCTION AND GENERAL IDEAS

It was suggested recently[1] that the methods and concepts of nonstandard analysis might be suitable for the investigation of systems with infinitely many degrees of freedom. As a concrete application of this suggestion, the existence and uniqueness of the thermodynamic limit was demonstrated for systems whose interaction potential satisfies a *strong* tempering condition. (See Sec. II, or Ref. 1, for definitions.) Although the use of nonstandard analysis allows a straightforward proof to be given for strongly tempered potentials,[1] it could be argued that this case is basically simple whatever method of proof is employed, so that the simplicity of this proof does not yet provide convincing evidence of the effectiveness of nonstandard analysis in physics. It is the purpose of the present paper to develop the suggestion further and show that the uniqueness and existence of the thermodynamic limit for weakly tempered systems and Coulomb systems can also be derived using nonstandard analysis. Furthermore, the logical structure of the proofs for these cases is exactly the same as that for the tempered case (treated in Ref. 1). The systems considered here are more complicated than the strongly tempered systems; the available proofs of the existence of the thermodynamic limit for these systems are a good deal more involved and demand more powerful analytical techniques than do the corresponding proofs for the tempered case.[2-4] It is, therefore, satisfactory that the nonstandard proofs proceed along the same pattern in all cases, although certain technical refinements are needed to handle the more complicated cases.

Nonstandard analysis is based on an ordered *field* $^*\mathbb{R}$, which is an *extension* of the field of real numbers $\mathbb{R}$. [For more details on nonstandard analysis, one could consult the book of Robinson,[5] Luxemburg[6] or, especially, Robinson and Lightstone.[7] Some explanations are contained in Ref. 1; an exposition of the field designed for physicists is in preparation (by Gambardella, Ostebee, and Dresden).] $^*\mathbb{R}$ contains, in addition to the real numbers, infinitesimals ϵ (ϵ is *smaller* than *any* arbitrarily small real number) and *infinite* numbers H (H is *larger* than *any* arbitrarily large real number). All algebraic operations and relations valid in $\mathbb{R}$ are equally valid in $^*\mathbb{R}$; for example, it is meaningful

to assert that $H_1 > H_2$, even if H_1 and H_2 are both infinite. With each function f defined on $\mathbb{R}$ is associated the natural extension $^*f$, which is defined on $^*\mathbb{R}$.[8] The main result used in the sequel is the transfer theorem. This theorem states (very roughly) that results established for a class of finite values of certain entities can be transferred to results valid for *infinite* values of these entities. It is this result which replaces the *process* of taking the limit in the usual treatment by the *substitution* of infinite values for appropriate variables in the present discussion.

Since there are a number of technical details, which might obscure the basically simple pattern of the proof, it might be helpful to outline the separate steps involved. The remainder of this section summarizes the qualitative features of the argument. The discussion is purposely kept on a heuristic level. Details and precise statements are contained in the succeeding sections. The basic object to be studied is the free energy per unit volume g,

$$g(\beta, \rho, \Lambda) \equiv \frac{1}{V(\Lambda)} \log Z(\beta, N, \Lambda). \tag{I.1}$$

In (I.1), Λ is a domain, $V(\Lambda)$ its volume, N is the number of particles in the system, Z is the canonical partition function, $\beta \equiv 1/kT$ is the inverse absolute temperature and ρ is the density defined by $\rho \equiv N/V(\Lambda)$. All quantities so far considered are *finite*. The interaction potential of the system will be called U. Different assumptions about U characterize different physical systems, but it is necessary to make a number of physical assumptions about the behavior of U before the system can reasonably be expected to exhibit thermodynamic behavior. These *assumptions* lead to *inequalities* for g, and a careful use of these inequalities leads, in turn, to the demonstration of the existence (and uniqueness) of the thermodynamic limit. Involved are the following steps:

(a) A condition on U is needed to eliminate a catastrophic collapse; the stability condition (see Refs. 2 and 3) is

$$U \geq -NE_0. \tag{I.2}$$

This condition gives rise to an upper bound for g,

$$g(\beta, \rho, \Lambda) < C(\rho, \beta) \quad \text{for all finite } \Lambda. \tag{I.3}$$

Since Λ is a finite volume, g is finite,

$$-\infty < g(\beta,\rho,\Lambda) < C(\rho,\beta) \text{ for all } \Lambda. \tag{I.4}$$

The shape of Λ does not enter condition (I.4) at all; it is valid for *all* finite volumes $V(\Lambda)$.

(b) It is further necessary to assume something about the long-range character of the potential. This is effectively an assumption about the forces between two clusters of particles, separated by a distance $R > R_0$; R_0 is a characteristic distance. From such "tempering" conditions it may be shown that other inequalities result,

$$g(\beta,\rho,\Lambda) \geqslant \sum \frac{V(\Omega_i)}{V(\Lambda)} g(\beta,\rho_i,\Omega_i) - \mathcal{R}. \tag{I.5}$$

In (5), Ω_i is a set of subdomains of Λ of volume $V(\Omega_i)$, $g(\beta,\rho_i,\Omega_i)$ is their local free energy, and $\mathcal{R}$ is a remainder term which *vanishes* for strongly tempered potentials. The analysis of (I.5) yields further inequalities necessary to demonstrate the thermodynamic limit. To use (I.5) effectively, the subdivision of Λ into sets Ω_i (and perhaps a remainder) has to be chosen judiciously. This introduces *packing* questions which deal with the way in which a domain Λ can be filled up (partially filled) with subdomains Ω_i of a particular type. It is at this point that the *type* of domain Λ and the nature of the subdomains Ω_i becomes of importance.

Both the standard and nonstandard proofs of the thermodynamic limit start from inequalities (I.4) and (I.5). (Different subdivisions Ω_i are used for the Coulomb, strongly and weakly tempered potentials.) In the standard proof, the inequalities are used to prove the convergence of a sequence of g values, calculated for suitably chosen sequences of volumes. These volumes Λ, must be packed in an appropriate way with subdomains Ω_i so that (5) can be applied. In the nonstandard method, by contrast, the transfer theorem is used to let Λ be an infinite element of *$\mathbb{R}$. It then remains to show that stg is *independent* of the particular infinite value Λ^* chosen. To apply the tempering condition (I.5), the volume Λ^* must be packed with appropriate domains Ω_i. Typically, the result obtained has the following form [See Ref. 1 for the definition of the standard part, as well as more details for the argument given after (I.7)]:

$$stg(\beta,\rho,\Lambda^*) \geqslant stg(\beta,\rho,\Omega) \text{ for all } \textit{finite cubes } \Omega. \tag{I.6}$$

In this case a volume Λ^* (of van Hove type) was packed by an infinite number of *finite* cubes Ω. Applying the transfer theorem and recalling that $stg(\beta,\rho,\Lambda^*)$ itself is *finite* for infinite Λ^* leads to the conclusion that

$$stg(\beta,\rho,\Lambda^*) \geqslant stg(\beta,\rho,\Omega^*) \tag{I.7}$$

for *all* cubes, be they finite or infinite. The generality of this result (obtained from the transfer theorem), allows an easy demonstration that $stg(\beta,\rho,\Lambda^*)$ is independent of Λ^* (whenever Λ^* is infinite), and thus the uniqueness of the thermodynamic limit.

From this brief outline, it is clear that the subdivision of Λ^* (where Λ^* is infinite) into subdomains, together with the appropriate packing theorems for infinite domains, are the main technical tools used. The scheme of the proof proceeds from conditions on U to inequalities for g when N and $V(\Lambda)$ are finite. The transfer theorem is used to obtain g for infinite volumes Λ^*. Next the transfer theorem is applied to the inequalities for g and to the packing theorems. These results are combined to show that stg becomes independent of Λ^*, when Λ^* is infinite. Of course, stg *does depend* on Λ, when Λ is finite.

The remainder of this paper is devoted to a detailed elaboration of the scheme outlined so far in general terms.

II. DEFINITIONS AND ASSUMPTIONS

In this section we collect the definitions and notations which will be used throughout this paper. When we consider systems which have hard cores the hard core radius will be denoted by δ and the maximum close packing density by ρ_c. In the case that there are no hard cores, $\delta = 0$ and $\rho_c = \infty$. For densities $\rho < \rho_c$ the free energy per unit volume is

$$g(\beta,\rho,\Lambda) \equiv \frac{1}{V(\Lambda)} \log Z(\beta,N,\Lambda), \tag{II.1}$$

where $Z(\beta,N,\Lambda)$ is the canonical partition function, Λ is a bounded open set in *$\mathbb{R}^\nu$ (a ν-dimensional Euclidean space), $V(\Lambda)$ is the volume of Λ Lebesgue measure), $\rho \equiv N/V(\Lambda)$, and β is the inverse temperature $\equiv 1/kT$. Classically,

$$Z(\beta,N,\Lambda) = \frac{\lambda^{-\nu N}}{N!} \int \cdots \int d^3 r_1 \cdots d^3 r_N \exp[-\beta U_N(\mathbf{r}_1 \cdots \mathbf{r}_N)] \tag{II.2}$$

where

$$\lambda^2 = \frac{h^2}{2\pi mkT}.$$

Quantum mechanically

$$Z(\beta,N,\Lambda) = \text{Tr}[\exp(-\beta H_N)], \tag{II.3}$$

$$H_N \equiv \sum_1^N \frac{\phi_i^2}{2m} + U_N(\mathbf{r}_1 \circ \circ \mathbf{r}_N). \tag{II.4}$$

Here $U_N(\mathbf{r}_1 \cdots \mathbf{r}_N)$ is the potential energy of the N particle configuration; $\mathbf{r}_i$ and ϕ_i are the particle positions and momenta respectively. Clearly, whenever $\rho < \rho_c$ and Λ is a finite volume, $g(\beta,\rho,\Lambda)$ will be finite. For a given domain Λ, the function $g(\beta,\rho,\Lambda)$ is defined only for densities which are *integral* multiples of $\rho_0 = 1/V(\Lambda)$. Following Fisher,[2] we extend the definition of $g(\rho,\beta,\Lambda)$ to arbitrary ρ by

$$g(\beta,\rho,\Lambda) = g(\beta,\rho',\Lambda,) + \eta[g(\beta,\rho' + \rho_0,\Lambda) - g(\beta,\rho',\Lambda)], \tag{II.5}$$

when

$$\rho = \rho' + \eta\rho_0, \quad \rho' = n\rho_0, \quad n \text{ is an integer}, \quad 0 \leqslant \eta < 1.$$

Using nonstandard analysis, we may extend the definition of $g(\beta,\rho,\Lambda)$ to domains Λ in *$\mathbb{R}^\nu$ which have an infinite volume. The problem of proving the existence of the thermodynamic limit of the free energy is to show that the standard function $G(\beta,\rho)$ defined by

$$G(\beta,\rho) \equiv stg(\beta,\rho,\Lambda) \tag{II.6}$$

for ρ in the half-open standard interval $[0,\rho_c)$ is the same for all *infinite* domains Λ. Thus, the problem is

to show $stg(\beta,\rho,\Lambda)=stg(\beta,\rho,\Lambda')$ for any pair of infinite domains Λ and Λ'. It is important to remember that the nonstandard extension of the standard function $G(\beta,\rho)$ is *not* $g(\beta,\rho,\Lambda)$. It should be noted further that Eq. (II.5) leads to the result that if $V(\Lambda)$ is infinite and ρ is *not* infinitesimally close to ρ_c, $g(\beta,\rho',\Lambda)\approx g(\beta,\rho,\Lambda)$ for all $\rho'\approx\rho$. This result plays an important role in the proofs presented in this paper. However, it does not imply that $G(\beta,\rho)$ is continuous.

We make the following physically sensible assumptions:

(A) $U_N(\mathbf{r}_1\cdots\mathbf{r}_N)$ is symmetric under the exchange of particle coordinates and translationally invariant.

(B) U satisfies

$$U_{N_1+N_2}(\mathbf{r}_1\cdots\mathbf{r}_{N_1},\mathbf{r}_1'\cdots\mathbf{r}_{N_2}')-U_{N_1}(\mathbf{r}_1\cdots\mathbf{r}_{N_1})$$
$$-U_{N_2}(\mathbf{r}_1'\cdots\mathbf{r}_{N_2}')\le N_1N_2W_B/R^{\nu+\epsilon}, \qquad \text{(II.7)}$$

whenever $|\mathbf{r}_i-\mathbf{r}_j'|\ge R>R_0$ for all $i=1,\ldots,N_1$, $j=1,\ldots,N_2$, and ϵ, R_0, and W_B are finite positive constants. The potential is strongly tempered if $W_B=0$, otherwise weakly tempered. This assures that the repulsive part of the interaction falls off sufficiently rapidly with distance, so that the system does not explode.

(C) It is assumed that the system (Hamiltonian) is H-stable. The Hamiltonian H_N is H-stable if

$$H_N\ge -NW_A. \qquad \text{(II.8)}$$

For classical systems it is sufficient that the *potential* U_N be *stable*; that is

$$U_N(\mathbf{r}_1\cdots\mathbf{r}_N)\ge -NW_A. \qquad \text{(II.8a)}$$

This requirement eliminates catastrophic interactions. If the Hamiltonian H_N is H-stable, then there exists a finite function $C(\beta,\rho)$ (independent of Λ) such that

$$g(\beta,\rho,\Lambda)\le C(\beta,\rho)\quad\text{for all }\Lambda. \qquad \text{(II.8b)}$$

This is the only place where H-stability is used in this paper. (II.8b) is a more precise version of the upper bound mentioned in Sec. I.

(D)[9] We assume that Λ is a bounded open set in $*\mathbb{R}^\nu$ such that $st\,V(\Lambda)/d(\Lambda)^\nu>0$, where $d(\Lambda)$ is the diameter of Λ. [The "diameter" $d(\Lambda)$ is the maximum distance between any two points on the boundary of Λ.]

(E) If $h\in *\mathbb{R}$ is such that $h/V(\Lambda)^{1/\nu}\approx 0$, then

$$V_h(\Lambda)/V(\Lambda)\approx 0. \qquad \text{(II.9)}$$

Here $V_h(\Lambda)$ is the volume of the set of points of Λ lying within a distance h of the boundary of Λ. Geometrically (E) asserts that "surface effects" should be negligible.

Together (D) and (E) are equivalent to the condition used by Fisher (see Appendix A) to prove the existence of the thermodynamic limit for systems with weakly tempered potentials. However, we feel that (E) is a much more natural requirement to impose.

If the potential U_N is strongly tempered, it is sufficient to require

(E*) If h is a *finite* standard real number, $V_h(\Lambda)/$

$V(\Lambda)\approx 0$. This condition is the nonstandard analog of van Hove convergence to infinity.

III. PACKING THEOREMS

In this section we collect several results on the packing of domains with cubes and balls. These theorems are purely geometrical in character and have *a priori* nothing to do with physics. However, the use of tempering inequalities such as I.6 requires results about the manner in which particles can be placed in subdomains; these in turn require packing theorems. We summarize one lemma, (due to Fisher), one packing theorem due to Lieb and Lebowitz, which are used in the proofs of two new packing theorems. These new results will be transcribed to infinite domains, since they are needed in that context. Theorem 1 is formulated directly in nonstandard terms; Theorem 3 can be directly transcribed (and is used) in the nonstandard framework.

Lemma 1 (Fisher): If a domain Ω is filled with cubes Γ of side d, lying entirely within Ω, the volume remaining when the maximum number of cubes has been inserted is *less* than $V_h(\Omega)$, where $h=\sqrt{\nu}\,d$ and V_h is defined in Sec. II.

The following packing theorem for infinite domains is needed when the potential U_N is weakly tempered. It allows us to fill *all* but an infinitesimal fraction of the volume of an *infinite* domain Λ with *infinite* cubes so that the distance between any two of the cubes is infinite.

Theorem 1: If Λ is a bounded open set in $*\mathbb{R}^\nu$, satisfying condition E, $V(\Lambda)$ is infinite and ϵ' is a standard real number such that $\epsilon/\nu(\nu+\epsilon)>\epsilon'>0$, then there exists a positive integer m (i.e., $m\in *\mathbb{N}^+$) and a positive real infinite number ω such that:

(i) m cubes Ω of edge $h=(1+1/\omega)^{-1}\,V(\Lambda)^{1/(\nu-\epsilon')}$ may be placed in Λ,

(ii) $mV(\Omega)/V(\Lambda)\approx 1$,

(iii) the m cubes may be placed in Λ so that they are separated from each other by a distance of at least $2R$, where $R=(1+\omega)^{-1}\,(V(\Lambda))^{1/(\nu-\epsilon')}$.

Proof: Choose $\alpha\approx 0$, so that $\alpha(V(\Lambda))^\tau$ is infinite; $\tau\equiv\epsilon/\nu-\epsilon_*(\nu+\epsilon)$.[10] Then

$$\omega\equiv[(\alpha V(\Lambda)^\tau)^{1/(\nu+\epsilon)}-1] \qquad \text{(III.1)}$$

is a positive infinite real number. Let $d=h+R=(V(\Lambda))^{1/(\nu-\epsilon')}$ and m be the maximum number of cubes Γ of edge d which can be placed in Λ so that they lie entirely within Λ. We can use Lemma 1 to obtain

$$1\ge\frac{mV(\Gamma)}{V(\Lambda)}>1-\frac{V_{h'}(\Lambda)}{V(\Lambda)}, \qquad \text{(III.2)}$$

where $h'=\nu^{1/2}d$. Since Λ was assumed to satisfy E, $V_{h'}(\Lambda)/V(\Lambda)\approx 0$ and so we must have

$$mV(\Gamma)/V(\Lambda)\approx 1. \qquad \text{(III.3)}$$

Centered in each of the cubes Γ we place a cube Ω of edge h. Clearly, the cubes Ω are separated by a distance of at least $2R$. Now

$$(h/d)=(1+1/\omega)^{-1}\approx 1, \qquad \text{(III.4)}$$

since ω is a positive infinite real number. Thus, $V(\Omega)/V(\Gamma) \approx 1$ and we have

$$m V(\Omega)/V(\Lambda) \approx 1. \tag{III.5}$$

This proves Theorem 1.

Theorem 2 describes a procedure (first given by Lieb and Lebowitz—Theorem 3.2) for the packing of a cube with balls such that the packing γ is exponentially fast.

Theorem 2 (Lieb and Lebowitz): Let $\alpha_\nu \equiv (2^\nu - 1)2\sqrt{\nu}$, σ_ν be the volume of an open ball of unit radius in R^ν, and $g_\nu \equiv 2^{-\nu}\sigma_\nu$. Let p be a positive integer and $1 + p \geq \alpha_\nu + g_\nu^{-1}$. Then it is possible to pack $\cup_{j=1}^{\infty}(n_j$ balls of radius $r_j)$ in an open ν-dimensional cube of volume σ_ν where $n_j \equiv p^{j-1}(1+p)^{j(\nu-1)}$ and $r_j \equiv (1+p)^{-j}$. The fraction of the volume filled after the balls of type j have been packed is $1 - \gamma^j$, where $\gamma = p/(1+p)$.

The meaning of the phrase "exponentially fast" is just that the fraction of volume left *unfilled* after the balls of type j have been packed, is $[(p/(p+1))]^j$ which decreases as a power with j.

If the potential U_N includes Coulomb interactions, the application of (1.5) requires yet another packing theorem. We need to pack a cube with a finite number of balls such that the particle number density in each ball is less than ρ_c. In addition, we require that the balls be separated from each other by at least R_0 where R_0 is defined in assumption (B). Theorem 3 establishes a sufficient condition on the size of a cube so that this is possible.

Theorem 3: If an integer k and positive real numbers $\bar{\rho}$ and r_0 satisfy the conditions

(a) $\rho_c > \bar{\rho} > \rho$,

(b) $r_0 > 2\nu^{1/2}(1/\delta - 2(\bar{\rho}))^{1/\nu} - 1$, \hfill (III.6)

(c) $r_0 > (\bar{\rho}|\sigma_\nu|)^{-1/\nu}$,

(d) $(\bar{\rho} - 1/r_0^\nu\sigma_\nu)(1 - \gamma^k)(1 + R/r_0)^{-\nu} > \rho$,

where R is a positive constant and δ is the hard core radius, then a cube Γ of edge $d = (R + r_0)(1+p)^k\sigma_\nu^{1/\nu}$ can be packed with $\cup_{j=1}^{k}(n_j$ balls of radius $R_j)$, where $R_j \equiv r_0(1+p)^{k-j}$ such that

(i) The distance between any two of the balls is at least $2R$ and the distance between a ball and the boundary of Γ is at least R;

(ii) At least $M = \rho V(\Gamma)$ particles can be accomodated in the balls without the density in any ball exceeding $\bar{\rho}$; and

(iii) Particles can be placed in the balls in such a way that the density $\hat{\rho}$ in any ball of type j satisfies the inequality

$$\rho' - 1/V_j \leq \hat{\rho} \leq \rho' + 1/V_j < \bar{\rho}, \tag{III.7a}$$

where

$$\rho' = \rho(1 - \gamma^k)^{-1}(1 + R/r_0)^\nu \quad \text{and} \quad V_j = R_j^\nu\sigma_\nu. \tag{III.7b}$$

Proof: Using Theorem 2 we pack the cube Γ with balls of radii $R_j' = (R + r_0)(1+p)^{k-j}$ where $j = 1, \ldots, k$. Concentric with each ball of radius R_j' we place a ball of radius R_j. Clearly, the distance between any two

balls of radii R_j and R_l $(1 \leq j, l \leq k)$ will be at least $2R$ and the distance between any ball of radius R_j $(1 \leq j \leq k)$ and the boundary of Γ will be at least R. This proves (i).

Lemma 1 implies that the maximum number of particles N_j' which may be placed in a ball of radius R_j satisfies the inequality

$$N_j' \geq \frac{(R_j - 2\delta\nu^{1/2})^\nu\sigma_\nu}{(2\delta)^\nu}. \tag{III.8}$$

Now,

$$\frac{\bar{\rho}V_j}{N_j} \leq \frac{\bar{\rho}(2\delta)^\nu}{(1 - 2\delta\nu^{1/2}/R_j)^\nu} \leq 1 \tag{III.9}$$

by condition (b) of the hypothesis. Since $[\bar{\rho}V_j] \leq \bar{\rho}V_j$, where $[\ ]$ denotes the greatest integer function, we conclude that at least $[\bar{\rho}V_j]$ particles may be placed in a ball of radius R_j. Let $\Omega \equiv \cup_{j=1}^{k}(n_j$ balls of $R_j)$ and M' the maximum number of particles which can be placed in Ω with the density in each *ball*, not exceeding $\bar{\rho}$. Using the result (III.9) we have

$$M' \geq \sum_{j=1}^{k} n_j[\bar{\rho}V_j]. \tag{III.10}$$

Condition (c) of the hypothesis requires that $\bar{\rho}V_k \geq 1$, so we can write

$$\bar{\rho}V_k = J + \xi, \tag{III.11}$$

where $J \geq 1$ is an integer and $0 \leq \xi < 1$. Since $V_j = V_k(1+p)^{(k-j)\nu}$, we have

$$[\bar{\rho}V_j] \geq (\bar{\rho}V_k - \xi)(1+p)^{(k-j)\nu}. \tag{III.12}$$

We use (12) to rewrite (10) as

$$M' \geq (\bar{\rho} - \xi/V_k) \sum_{j=1}^{k} n_j V_k(1+p)^{(k-j)\nu}$$

$$= (\bar{\rho} - \xi/V_k) \sum_{j=1}^{k} n_j V_j. \tag{III.13}$$

From Theorem 1 we have

$$\sum_{j=1}^{k} n_j V_j = (1 - \gamma^k)\bar{V}, \tag{III.14}$$

where $\bar{V} = [r_0(1+p)^k]^\nu$.

Now

$$\frac{\bar{V}}{V(\Gamma)} = \left(\frac{r_0}{R + r_0}\right)^\nu, \tag{III.15}$$

so that (14), (15), and condition (d) of the hypothesis may be used to obtain

$$M' > \rho V(\Gamma) = M. \tag{III.16}$$

This proves (ii).

Condition (d) of the hypothesis gives

$$\bar{\rho} > \rho' + 1/V_k > \rho' + 1/V_j \quad \text{for } j < k. \tag{III.17}$$

We recall that

$$\rho' \sum_{j=1}^{k} n_j V_j = \rho V(\Gamma) = M, \tag{III.18}$$

so that

$$\sum_{j=1}^{k} n_j(\rho' - 1/V_j)V_j \leq M \leq \sum_{j=1}^{k} n_j(\rho' + 1/V_j)V_j. \qquad \text{(III.19)}$$

This proves (iii) and the theorem.

After exhibiting these geometrical theorems, we return to the physical situations, where these results will be used.

IV. WEAKLY TEMPERED POTENTIALS

In this section we present a proof of the existence of the thermodynamic limit for a classical system with an interaction potential satisfying conditions (A)—(C). The arguments remain essentially unchanged for quantum systems.

The assumption (B) may be used to derive the inequality[2]

$$g(\beta, \rho, \Lambda) \geq \sum_{i=1}^{n} \frac{V(\Omega_i)}{V(\Lambda)} g(\beta, \rho_i, \Omega_i) - \beta W_B \rho^2 \frac{V(\Lambda)}{R^{\nu+\epsilon}}, \qquad \text{(IV.1)}$$

where $\cup_{i=1}^{n} \Omega_i \subset \Lambda$, $\Omega_i \cap \Omega_j = \phi$, and the subdomains Ω_i are separated from each other by a distance $R > R_0$. Further, $\rho V(\Lambda) = \sum_{i=1}^{n} \rho_i V(\Omega_i)$.

This inequality (IV.1) and the upper bound (II.10) may be used to prove the existence of the thermodynamic limit for cubic domains. The proof of this statement, using nonstandard analysis is *not* very different from the proof given by Fisher. We *assume* this result, without casting it in the language of nonstandard analysis. The proof of the existence of the thermodynamic limit using nonstandard analysis for *general* domains is quite different from Fisher's proof. For that reason it is presented here while the proof for cubic domains is omitted.

Let Λ be an infinite volume bounded open set in $^*R^\nu$ satisfying assumptions (D) and (E). Choose $\epsilon' \in \mathbb{R}$ to satisfy the inequality $\epsilon/\nu(\nu + \epsilon) > \epsilon' > 0$. We may apply the inequality (IV.1) to the packing of Λ^* with cubes Ω as described in Theorem 1,

$$g(\beta, \rho, \Lambda^*) \geq \frac{m V(\Omega)}{V(\Lambda^*)} g(\beta, \rho', \Omega) - \beta W_B \rho^2 \frac{V(\Lambda^*)}{R^{\nu+\epsilon}}, \qquad \text{(IV.2)}$$

where

$\rho' = V(\Lambda^*)/m V(\Omega)\rho \approx \rho$ and $R > R_0$ is defined in

$$\text{Theorem 1.} \qquad \text{(IV.2')}$$

The existence of the thermodynamic limit of the free energy for cubic domains and the fact that $V(\Lambda^*)/R^{\nu+\epsilon} = \alpha \approx 0$ (α is defined in the proof of Theorem 1) imply that $g(\beta, \rho, \Lambda^*)$ is bounded *below* by a finite number. Since a finite upper bound for $g(\beta, \rho, \Lambda^*)$ exists by virtue of the bound (II.8b) we may take the standard part of (IV.1). This results in

$$st\,g(\beta, \rho, \Lambda^*) \geq st\,g(\beta, \rho', \Omega) = st\,g(\beta, \rho, \Omega). \qquad \text{(IV.3)}$$

The last equality follows since $\rho' \approx \rho$. Let Ω^* be an infinite cube enclosing Λ^*, such that $st\,(V(\Lambda^*)/V(\Omega^*)) = \mu > 0$. The existence of such a cube follows from condition (D). Let α' be an infinitesimal chosen such that $\omega' = [\alpha' V(\Omega^*)^{\epsilon/\nu}]^{1/(\nu+\epsilon)}$ is infinite. Define $R_1 = V(\Omega^*)^{1/\nu}/\omega'$ and let Γ^* denote the set of points of $\Omega^* \backslash \Lambda^*$ (the complement of Λ^* in Ω^*) lying at least a distance R_1

from the boundary of Λ^*. We place the particles in the domains Λ^* and Γ^* in such a way that the density in Λ^* is ρ and the density in Γ^* is $\rho' = (V(\Omega^*) - V(\Lambda^*))/V(\Gamma^*)\rho \approx \rho$.

Since $R_1 > R_0$ we can apply the inequality (IV.1) to the subdomains Λ^* and Γ^* of Ω^*,

$$g(\beta, \rho, \Omega^*) \geq \frac{V(\Lambda^*)}{V(\Omega^*)} g(\beta, \rho, \Lambda^*) + \frac{V(\Gamma^*)}{V(\Omega^*)} g(\beta, \rho', \Gamma^*)$$

$$- \beta W_B \rho^2 \frac{V(\Omega^*)}{R_1^{\nu+\epsilon}}. \qquad \text{(IV.4)}$$

Now $V(\Omega^*)/R_1^{\nu+\epsilon} = \alpha' \approx 0$, so that we have

$$st\,g(\beta, \rho, \Omega^*) \geq \mu\,st\,g(\beta, \rho, \Lambda^*) + (1 - \mu)\,st\,g(\beta, \rho', \Gamma^*). \qquad \text{(IV.5)}$$

The result (IV.3) holds for any infinite domain Λ^* satisfying (D) and (E); it does *not* depend on Ω^* being a subdomain of E. Since Γ^* satisfies (D) and (E) we can write

$$st\,g(\beta, \rho, \Omega^*) \geq \mu\,st\,g(\beta, \rho, \Lambda^*) + (1 - \mu)\,st\,g(\beta, \rho', \Omega^*). \qquad \text{(IV.6)}$$

This implies directly that

$$st\,g(\beta, \rho, \Omega^*) \geq st\,g(\beta, \rho, \Lambda^*). \qquad \text{(IV.7)}$$

Equations (IV.7) and (IV.3) yield the main result,

$$st\,g(\beta, \rho, \Omega^*) = st\,g(\beta, \rho, \Lambda^*). \qquad \text{(IV.8)}$$

Since (IV.8) holds for *all* infinite domains satisfying (D) and (E), this proves the existence of the thermodynamic limit for general domains Λ^*. This completes the proof of the thermodynamic limit for weakly tempered case. It is noteworthy that once the packing theorem is established, the remainder of the proof is identical with that in the strong tempering case.

V. NEUTRAL COULOMB SYSTEMS

The Coulomb interaction is not even weakly tempered, so a different proof of the existence of the thermodynamic limit is needed. If all the particles in the system have charges of the same sign, the Hamiltonian of the system is H-stable, but obviously there is *no* thermodynamic behavior. We, therefore, consider only systems which are *overall* neutral. For these systems we can anticipate thermodynamic behavior whenever the Hamiltonian is H-stable. If, in addition to the Coulomb interaction, there is a tempered interaction which includes a hard core, the Hamiltonian *will* be H-stable, in both classical[11] and quantum mechanics.[12] Dyson and Lenard[13] and, more recently, Federbush[14] and Lieb and Thirring,[15] have shown that the purely Coulomb Hamiltonian is H-stable in three dimensions if all the positively charged particles, and/or all the negatively charged particles, are fermions.

We now restrict our considerations to overall neutral systems in three dimensions ($\nu = 3$) and assume that all the negatively charged particles are fermions (i.e., we deal with *real* matter). In addition to a Coulomb interaction, a rotationally invariant interaction satisfying (A)—(C) may be present.

The method of proof used here may be easily adapted

to the case of a classical system (e. g. , a neutral system of charged particles with hard cores).

A. Neutral Coulomb case and weak tempering

It is advantageous to give separate proofs for the two cases where the non-Coulomb part of the potential is strongly and weakly tempered. In this section, we will discuss the second case. When describing particles (or mixtures of particles having different charges) and densities we will use, throughout, the language introduced by Lieb and Lebowitz[2,4] of a fundamental multiplet of particles, which is overall electrically neutral. Densities and particle numbers are understood to be given in terms of this multiplet as a basic unit. In principle, the density ρ should be treated as a vector if an arbitrary mixture of particles is considered. Since the main purpose of this system is to demonstrate the use of nonstandard analysis, ρ is treated as a scalar. It takes only minor (mainly notational) modifications to generalize the results to arbitrary—but overall neutral—mixture of charged particles. The basic outline of the proof is again the same: Using the H-stability of the potentials and the tempering inequalities, bounds are established for g. Then the transfer theorem is applied, eventually leading to the independence of $stg(\Lambda^*)$ on Λ^*. In the present Coulomb case, a basic tempering inequality (analogous) to (IV. 1) was established by Lieb and Lebowitz[4] (Theorem 2. 6). Adapted to the present case it reads:

$$g(\beta,\rho,\Omega) \geq \sum_{j=1}^{k} \frac{V(B_j)}{V(\Omega)} g(\beta,\rho_j,B_j) + \frac{V(\Lambda)}{V(\Omega)} g(\beta,\rho_\Lambda,\Lambda)$$

$$- \beta\rho^2 W_B \frac{V(\Omega)}{R^{\nu+\epsilon}} . \qquad (V.1)$$

In (V.1), Ω is the domain under consideration, Λ is a subdomain of Ω, and the B_j are disjoint open balls which are subdomains of Ω/Λ, i. e. ,

$$\Omega \supset \bigcup_{j=1}^{k} B_j \cup \Lambda.$$

Further, the B_j are separated from each other and from Λ by a distance larger than R. Finally, the *net* charge in each ball B_j is zero, and the local densities ρ_j and ρ_Λ satisfy

$$\rho V(\Omega) = \sum_{j=1}^{k} \rho_j V(B_j) + \rho_\Lambda V(\Lambda). \qquad (V.2)$$

The conditions on the domains together with the properties of the Coulomb potential enable us (Lieb and Lebowitz!) to obtain an inequality which has the appearance of a tempering inequality even though the Coulomb potential has no tempering properties. To apply the inequality (V.1), special packing procedures are necessary and special domains will be selected. This is why the packing theorems—especially Theorem 3—are so essential. We will present the proof of the thermodynamic limit for arbitrary domains Ω, satisfying (D) and (E), *assuming* the result for ball domains (as given by Lieb and Lebowitz). The derivation of the result for general domains is simplified most by the use of nonstandard analysis. (It will, of course, be possible to show the thermodynamic limit for ball domains, using nonstandard analysis, but that would en-

tail little more than a repetition of the Lieb and Lebowitz argument, combined with the method of Ref. 1, so it can be omitted.)

Since g has an upper bound for all finite Λ (from the stability condition), $g(\Lambda^*)$ is finite even when $V(\Lambda^*)$ satisfies the requirements (D) and (E). Pack Λ^* first with cubes Γ of infinite edge. Next, use Theorem 3 (Sec. III) to pack the cubes Γ with balls of infinite radius. It is clear that the nonstandard version of Theorem 3 is presupposed in this discussion. We then apply inequality (V.1) to obtain a lower bound on the free energy $g(\beta,\rho,\Lambda)$ in terms of the free energy of an infinite ball. We then use the same method to obtain an upper bound on $g(\beta,\rho,\Lambda)$. Together, these bounds may be used to establish that stg is *independent* of Λ, which demonstrates the uniqueness of the thermodynamic limit.

Let ϵ' be a standard real number chosen so that $\epsilon/\nu(\nu+\epsilon) > \epsilon' > 0$. Also let α be an infinitesimal real number, and k an infinite integer chosen such that Δ is infinite (p is a finite positive integer),

$$\Delta \equiv \alpha(V(\Lambda))^{\epsilon/\nu(\nu+\epsilon)-\epsilon'}(1+p)^{-k}. \qquad (V.3)$$

Define

$$R \equiv \alpha(1+p)^{-k}\sigma_\nu^{-1/\nu}(V(\Lambda))^{1/\nu-\epsilon'}, \qquad (V.4a)$$

$$r_0 \equiv (1-\alpha)(1+p)^{-k}\sigma_\nu^{-1/\nu}(V(\Lambda))^{1/\nu-\epsilon'}. \qquad (V.4b)$$

Notice that both R and r_0 are infinite. (σ_ν was defined in Sec. III, Theorem 2.) If $\bar{\rho}$ is chosen as a *standard* real number such that $\rho_c > \bar{\rho} > \rho$, then the numbers $\bar{\rho}$, r_0, R, and k satisfy the hypothesis of the nonstandard extension of Theorem 3 (Sec. III). The cubes Γ of Theorem 3 are of edge $d = (V(\Lambda))^{1/\nu-\epsilon'}$. Since $d/V(\Lambda)^{1/\nu} \approx 0$, there is an integer $m \in {}^*\mathbf{N}^+$ such that m cubes Γ may be placed in Λ and

$$mV(\Gamma)/V(\Lambda) \approx 1. \qquad (V.5)$$

(See the proof of Theorem 1.) Define M and θ by

$$N/m = M + \theta/m, \quad 0 \leq \theta < m, \quad M, \theta \in {}^*\mathbf{N}^+. \qquad (V.6)$$

Here $N = \rho V(\Lambda)$. Now place $(M+1)$ particles in each of θ of the cubes Γ, and M particles in each of the rest. Then the particle density in each of the cubes Γ is less than $\bar{\rho}$ (by construction) because

$$\frac{M+1}{V(\Gamma)} \leq \frac{V(\Lambda)}{mV(\Gamma)}\rho + \frac{1}{V(\Gamma)} \approx \rho. \qquad (V.7)$$

(By assumption, $\bar{\rho}$ is a standard real number greater than ρ and thus greater than any real number infinitesimally close to ρ.) We, therefore, may apply the inequality (V.1) to the packing of cubes Γ described in Theorem 3 to obtain

$$g(\beta,\rho,\Lambda) \geq \frac{mV(\Gamma)}{V(\Lambda)} \sum_{j=1}^{k} \frac{n_j V(B_j)}{V(\Lambda)} g(\beta,\rho_j,B_j)$$

$$- \beta\rho^2 W_B \frac{V(\Lambda)}{R^{\nu+\epsilon}} , \qquad (V.8)$$

where B_j denotes an open ball of radius R_j. The balls B_j ($j = 1,\ldots,k$) are all finite so $\rho_j \approx \rho$ by Theorem 3. From the existence of the thermodynamic limit for ball domains, we conclude that

$$g(\beta, \rho_j, B_j) = st\,(g(\beta, \rho, B_0)) + \eta_j, \qquad \text{(V.9a)}$$

$$\eta_j \approx 0. \qquad \text{(V.9b)}$$

Here B_0 is any one of the infinite balls; the free energy of the different infinite balls can only be infinitesimally different. Use (V.9) in (V.8) to obtain

$$g(\beta, \rho, \Lambda) \geq \frac{mV(\Gamma)}{V(\Lambda)}\left[st\,g(\beta, \rho, B_0) + \eta'\right]\sum_{j=1}^{k}\frac{n_j V(B_j)}{V(\Gamma)}$$

$$\qquad - \beta\rho^2 W_B \frac{V(\Lambda)}{R^{\nu+\epsilon}}$$

$$= \frac{mV(\Gamma)}{V(\Lambda)}\left[st\,g(\beta, \rho, B_0) + \eta'\right](1 - \gamma^k) - \beta\rho^2 W_B \frac{V(\Lambda)}{R^{\nu+\epsilon}}$$

$$\qquad\qquad\qquad\qquad\qquad\qquad\qquad \text{(V.10)}$$

where $\eta' = \mathrm{Min}_j\,\eta_j \approx 0$, and γ was defined in Sec. II. Now $V(\Lambda)/R^{\nu+\epsilon} = (\Delta\sigma_\nu^{-1/\nu})^{-(\nu+\epsilon)} \approx 0$, so that using (V.5), and that k is an infinite integer we have

$$st\,g(\beta, \rho, \Lambda) \geq st\,g(\beta, \rho, B_0). \qquad \text{(V.11)}$$

Next let Ω be an infinite ball enclosing the domain Λ, such that $st\,(V(\Lambda)/V(\Omega)) = \mu > 0$. The existence of such a ball follows from condition (D). Let Ω' denote the set of points of $\Omega \setminus \Lambda$ at least a distance R from the boundary of Λ. Since Ω and Λ satisfy the requirements (D) and (E), so does Ω'. We place the particles in Ω' and Λ in such a way that the *net* charge in Λ is zero, and the density in Λ is ρ. The density in Ω' is then ρ' $= (V(\Omega) - V(\Lambda))/V(\Omega')\rho \approx \rho$. Since $st\,(V(\Omega')/V(\Omega)) = 1 - \mu$, we have $d/(V(\Omega'))^{1/\nu} \approx 0$ so that we can pack Ω' with the same cubes Γ used before. That is, there is an $m' \in {}^*\mathbf{N}^+$ such that m' cubes Γ can be placed in Ω' and

$$m'V(\Gamma)/V(\Omega') \approx 1. \qquad \text{(V.12)}$$

The arguments which led to (V.8) may be repeated to derive the inequality

$$g(\beta, \rho, \Omega) \geq \frac{V(\Lambda)}{V(\Omega)}g(\beta, \rho, \Lambda) + \frac{m'V(\Gamma)}{V(\Omega')}\sum_{j=1}^{k}\frac{n_j V(B_j)}{V(\Gamma)}g(\beta, \rho_j, B_j)$$

$$\qquad - \beta\rho^2 W_B \frac{V(\Omega)}{R^{\nu+\epsilon}} \qquad \text{(V.13)}$$

where

$$\rho_j \approx \rho' \quad \text{for } j = 1, \ldots, k.$$

The arguments which led from (V.8) to (V.11) can now be applied to (V.13) and yield

$$st\,g(\beta, \rho, B_0) \geq st\,g(\beta, \rho, \Lambda). \qquad \text{(V.14)}$$

Equations (V.11) and (V.14) clearly imply that $st\,g(\beta, \rho, B_0) = st\,g(\beta, \rho, \Lambda)$ for any Λ satisfying the requirements (D) and (E). This establishes the independence of $st\,g$ on the infinite domain Λ, hence the uniqueness of the thermodynamic limit for the weakly tempered Coulomb case.

B. Neutral Coulomb case and strong tempering

In this section we sketch, rather briefly, the modifications necessary in the proof, in the case that the additional potential is strongly rather than weakly tempered. It is, of course, true that strong tempering implies weak tempering, so in a sense this proof is not necessary. However, it is interesting to see that the strong tempering condition allows the demonstration of

the thermodynamic limit for more general domains [satisfying condition (E*) rather than (E)]. Thus the tempering conditions influence the *type* of domain for which the thermodynamic limit can be shown to exist.

Consider again an infinite domain Λ of volume $V(\Lambda)$ and density $\rho = N/V(\Lambda)$. The method of proof is similar to that employed in the neutral Coulomb and weak tempering case, with the *main* difference that we must now pack the infinite domain Λ with *finite* rather than infinite cubes. Choose $\bar{\rho}$, r_0, and k to satisfy the conditions of Theorem 3 (Sec. III). Let $\bar{R} = (r_0 + R_0)$ $\times (1 + \rho)^k \sigma_\nu^{1/\nu}$, where R_0 is the "strong tempering distance" defined in Sec. I. Further, whenever R appears in Theorem 3 it has been replaced by R_0. We can pack Λ with m (which is infinite) finite cubes Γ of edge $\bar{R}$ so that

$$mV(\Omega)/V(\Gamma) \approx 1. \qquad \text{(V.15)}$$

Define M and θ by the relations

$$N/m = m + \theta/M, \qquad 0 \leq \theta < m. \qquad \text{(V.16)}$$

Now place $(M + 1)$ particles in θ of the cubes and M particles in the rest.

Let $\rho' \equiv (V(\Lambda)/mV(\Gamma))\rho \approx \rho$. Then

$$\frac{M+1}{V(\Gamma)} \leq \rho' + \frac{1}{V(\Gamma)} \leq \rho' + \frac{1}{V_k} < \bar{\rho}. \qquad \text{(V.17)}$$

In (V.17), V_k is the volume of the *ball* with smallest radius, and condition (d) of Theorem 3 (Sec. III) has been used. We now apply Theorem 3 to the cubes Γ, i.e., each Γ is packed with balls B_j, satisfying (i), (ii), and (iii) of Theorem 3. Applying the fundamental inequality (V.1) to this subdivision leads to

$$g(\beta, \rho, \Lambda) \geq \frac{mV(\Gamma)}{V(\Lambda)}\sum_{j=1}^{k}\frac{n_j V(B_j)}{V(\Gamma)}g(\beta, \rho_j, B_j). \qquad \text{(V.18)}$$

The ρ_j are selected as in Theorem 3.

Notice, parenthetically, that a "mechanical naive neglect" of the last term in (V.8) would just yield (V.18). However, in (V.18) we know that all the B_j are finite [something a naive use of (V.8) could not establish], since further $\rho_j < \rho_c$ for all j, $g(\beta, \rho_j, B_j)$ is a finite standard number for all j. In addition, $g(\beta, \rho, \Lambda)$ has a finite upper bound (from H-stability). Therefore, we can take the standard part of (V.18), giving

$$st\,g(\beta, \rho, \Lambda) \geq \sum_{j=1}^{k}n_j\frac{V(B_j)}{V(\Gamma)}g(\beta, \rho_j, B_j). \qquad \text{(V.19)}$$

Equation (V.19) holds for any r_0 and k satisfying the conditions of Theorem 3. If r_0 and k satisfy these conditions, so do all r_0' and k', satisfying $r_0' > r_0$ and $k' > k$. We can, therefore, apply the transfer theorem to (V.19), which establishes the validity of (V.19) for all infinite values of $r_0 \in {}^*\mathbf{R}$ and $k \in {}^*\mathbf{N}^+$. Thus Equation (V.19) also applies to infinite balls B_j; then by Theorem 3, $\rho_j \approx \rho$. The arguments following Eqs. (V.9)–(V.11) can just be repeated to yield the result

$$st\,g(\beta, \rho, \Lambda) \geq st\,g(\beta, \rho, B_0). \qquad \text{(V.20)}$$

The remainder of the proof is pretty much a repetition of that given in Sec. I Part A. Let Ω be a ball enclosing Λ, such that $st(V(\Lambda)/V(\Omega)) = \mu > 0$. ($\mu$ is a standard real number.) Let Ω denote the set of points of Ω

exterior to Λ, which are a distance greater than R_0 from the boundary of Λ. Since Ω' satisfies (E*), we can pack it with *finite* cubes of side $\bar{R}$. As before, we pack these cubes with balls B_j. Applying the inequality (V. 1) to this subdivision of the ball Ω, yields an inequality similar to (V. 13),

$$g(\beta, \rho, \Omega) \geq \frac{V(\Lambda)}{V(\Omega)} g(\beta, \rho, \Lambda) + \frac{m V(\Gamma)}{V(\Omega')} \left(\frac{V(\Omega')}{V(\Omega)} \right)$$

$$\times \sum_{j=1}^{k} \frac{n_j V(B_j)}{V(\Gamma)} g(\beta, \rho_j, B_j). \tag{V. 21}$$

The argument proceeds now as before [Eqs. (V. 8)– (V. 11)]; to avoid repetitious detail, just the result is recorded,

$$stg(\beta, \rho, B_0) \geq stg(\beta, \rho, \Lambda). \tag{V. 22}$$

Equations (V. 20) and (V. 22) imply the existence and uniqueness of the thermodynamic limit for neutral Coulomb strongly tempered systems with the conditions (D) and (E*) on the domains.

VI. CONCLUSIONS

The proofs presented in this paper would seem to be reasonably direct demonstrations of the existence and uniqueness of the thermodynamic limit for the weakly tempered and Coulomb systems. The physical input required consists of inequalities satisfied by the free energy as a consequence of the stability and tempering conditions of the interaction potentials. These results were taken from the basic studies of Fisher, Ruelle, Lieb, and Lebowitz. However, from these results, with the use of the packing theorems, repeated application of the transfer theorem of nonstandard analysis leads in a direct way to the existence and uniqueness of the thermodynamic limit. Comparison with the work of Fisher, Ruelle, Lieb, and Lebowitz shows that the nonstandard method avoids the limiting process $(N \rightarrow \infty, V \rightarrow \infty)$ which is usually both difficult and delicate. It is pertinent to observe that this limiting process becomes more involved as the systems and domains considered become more complicated. By contrast, the nonstandard method of proof remains much the same for different systems and domains. The tempering inequalities—which are needed in both approaches—become more involved, but the main complication of, say, the weakly tempered, compared to the strongly tempered, or the Coulomb proof, is the appropriate subdivision of the domain to which the tempering inequalities are to be applied. The needed packing theorems for different domain are the expression of this additional complication, but only need to be applied once. With the results of this paper, we believe that we have demonstrated the effectiveness of nonstandard analysis for questions dealing with the thermodynamic limit. It should be stressed that so far we have reproduced *known* results, albeit with different and, we believe, more straightforward methods. It would seem that all the known results on the thermodynamic limit can be obtained in this manner. Much more interesting is, of course, the question of whether *new* physical results can be obtained with these methods. We have some partial results, pertaining to the thermodynamic limit of correlation functions and the

uniqueness of the equilibrium state (which is the state reached when the time $t \rightarrow \infty$). It appears that the greater the part of the analysis that can actually be carried out in *$\mathbb{R}$, the more effective the use of nonstandard analysis becomes. It takes some time and some effort to become familiar with that language, but it seems fair to conclude that nonstandard analysis is a sufficiently promising method to continue its further exploration. The results on the correlation functions and the uniqueness of the equilibrium state will be reported in a future publication.

APPENDIX

Fisher[2] makes the following assumptions about the domains Λ, considered in the proof of the existence of the thermodynamic limit of the free energy, for systems interacting via weakly tempered potentials:

(i) Λ is a connected, bounded open set in *$\mathbb{R}^\nu$;

(ii) $\lim_{h \rightarrow 0} V_h(\Lambda)/V(\Lambda) = 0$;

(iii) there exists a *shape* function $\pi(\alpha)$ satisfying

    (iiia) $\lim_{\alpha \rightarrow 0} \pi(\alpha) = 0$,

    (iiib) there is a finite constant α', so that for $\alpha < \alpha'$,

$$V_h(\Lambda)/V(\Lambda) \leq \pi(\alpha),$$

where $h = \alpha V(\Lambda)^{1/\nu}$.

In this paper, conditions (ii) and (iii) were *not* assumed. Other conditions were encountered in the context of our proof, which at first sight appeared more natural,

(ii)' if $h/(V(\Lambda))^{1/\nu} \approx 0$, then $V_h(\Lambda)/V(\Lambda) \approx 0$;

(iii)' if $d(\Lambda)$ is the diameter of Λ,

$$\mu = st \left(\frac{V(\Lambda)}{d(\Lambda)^\nu} \right) > 0.$$

(ii)' and (iii)' have the advantage that no shape function need to be introduced. (ii)' asserts that for infinitesimal "skins" the volume of the skin is infinitesimal, which is a reasonable condition. (iii)' gives some condition on the shape of the volume.

It is the purpose of this Appendix to prove that the conditions (ii)' and (iii)' are, in fact, equivalent to the conditions (ii) and (iii) of Fisher. Condition (ii)' clearly implies (ii). We now show that (ii)' and (iii)' imply the condition (iii). We have from (iii)' that

$$1 \geq V(\Lambda)/(d(\Lambda))^\nu \geq \mu. \tag{A1}$$

If a cube of edge $d(\Lambda)$ enclosing Λ is placed in the center of a cube of edge $2d(\Lambda)$, we see that

$$1 \geq \frac{V(\Lambda)}{2^\nu (d(\Omega))^\nu} \geq \frac{\mu}{2^\nu}. \tag{A2}$$

Further, for any $h < \frac{1}{2} d(\Lambda)$,

$$V_h(\Lambda) \leq 2^\nu (d(\Lambda))^\nu - V(\Lambda). \tag{A3}$$

From (A2) and (A3) we obtain

$$\frac{V_h(\Lambda)}{V(\Lambda)} \leq \frac{2^\nu (d(\Lambda))^\nu}{V(\Lambda)} - 1 \leq \frac{2^\nu}{\mu} - 1. \tag{A4}$$

Clearly $2^\nu/\mu - 1$ is finite and positive [see (A2)]. Now consider the function $\pi(\alpha)$ defined by

$$\pi(\alpha) \equiv V_h(\Lambda)/V(\Lambda), \qquad (A5)$$

where

$$h = \alpha(V(\Lambda))^{1/\nu}. \qquad (A5')$$

It is clear that $\pi(\alpha)$ satisfies condition (iii) of Fisher for $\alpha < \alpha' < d(\Lambda)/2(V(\Lambda))^{1/\nu}$, since the domain Λ was assumed to satisfy condition (ii)'. Thus, (ii)') and (iii)' imply (ii) and (iii).

Fisher proves that (iii) implies that Λ can be enclosed in a cube Γ, with $V(\Lambda)/V(\Gamma) \geqslant \mu > 0$, where μ is a standard real number, which is precisely condition (iii)'. Assumption (ii) also implies (ii)'. We have shown here that (ii)' and (iii)' imply (ii) and (iii) of Fisher, while Fisher showed that (ii) and (iii) imply (ii)' and (iii)'. Hence, the equivalence is established.

[1]A. Ostebee, P. Gambardella, and M. Dresden, Phys. Rev. A **13** (1976).

[2]M.E. Fisher, Arch. Rat. Mech. Anal. **17**, 377 (1964).

[3]D. Ruelle, Helv. Phys. Acta **36**, 189 (1963).

[4]E. Lieb and J. Lebowitz, Adv. Math. **9**, 316 (1972).

[5]A. Robinson, *Non-standard Analysis* (North-Holland, Amsterdam, 1970).

[6]W.A.J. Luxemburg, *Non-standard Analysis* (California Institute of Technology, Pasadena, Calif., 1962).

[7]A.H. Lightstone and A. Robinson, *Non-Archimedian Fields and Asymptotic Expansions* (North-Holland, Amsterdam, 1975).

[8]The notation used is not always consistent. In many ways it would be best to denote entities in $^*\mathbb{R}$, with a star such as $^*\epsilon$, $^*H$. In $^*\mathbb{R}$, entities can be finite or infinite. A function on $^*\mathbb{R}$ should be written as $^*f$. This is not always done—one might get * crazy. Usually when the transfer theorem is applied, necessary stars are indicated in a few of the succeeding formulas. Once the context is understood, the stars are often omitted again.

[9]In the first paper, where *strong* tempering was assumed throughout, it was sufficient to require a condition on the domains weaker than D.

[10]This is possible because any infinite number may be written as the product of two other infinite numbers. If A, B, C, are infinite real numbers and $A = BC$, then $AB^{-1} = C$ and $B^{-1} \simeq 0$.

[11]L. Onsager, J. Chem. Phys. **43**, 189 (1939).

[12]M.E. Fisher and D. Ruelle, J. Math. Phys. **7**, 260 (1966).

[13]F. Dyson and A. Lenard, J. Math. Phys. **8**, 423 (1967); **9**, 698 (1968).

[14]P. Federbush, J. Math. Phys. **16**, 347, 706 (1975).

[15]E. Lieb and W. Thirring, Phys. Rev. Lett. **35**, 687 (1975).

Thermodynamics

Frank J. Zeleznik

Lewis Research Center, NASA, Cleveland, Ohio 44135

The fundamental structure of thermodynamics is purely algebraic, in the sense of atopological, and it is also independent of partitions, composite systems, the zeroth law, and entropy. The algebraic structure requires the notion of heat, but not the first law. It contains a precise definition of entropy and identifies it as a purely mathematical concept. It also permits the construction of an entropy function from heat measurements alone when appropriate conditions are satisfied. Topology is required only for a discussion of the continuity of thermodynamic properties, and then the weak topology is the relevant topology. The integrability of the differential form of the first law can be examined independently of Caratheodory's theorem and his inaccessibility axiom. Criteria are established by which one can determine when an integrating factor can be made intensive and the pseudopotential extensive and also an entropy. Finally, a realization of the first law is constructed which is suitable for all systems whether they are solids or fluids, whether they do or do not exhibit chemical reactions, and whether electromagnetic fields are or are not present.

INTRODUCTION

Thermodynamics, perhaps more than any other physical theory, can be said to pervade the fabric of science. Its history is measured in centuries rather than decades. Its range of applicability appears to be limitless. Its fundamental principles seem to be almost trivially simple. Its practitioners, both contemporary and historical, are legion. Yet in spite of all of these things, many students and practitioners of thermodynamics feel a vague, and sometimes not so vague,[1] uneasiness about the subject and its fundamentals. Much of this malaise can be ascribed to the second law of thermodynamics and the concept of entropy because the zeroth and first laws seem to be quite readily accepted.

Heat engines, partitions of various kinds, and composite systems form the substance from which thermodynamic theory is generally constructed. Thus, for example, heat engines are the devices which were used in the early formulations of the entropy concept, and their use for this purpose has persisted to the present. Partitions are used to control the transfer of heat and various forms of work between a system and its surroundings or among subsystems forming a composite system. If thermodynamics is viewed solely as the study of heat, work, and their interconversion, then this is a perfectly acceptable procedure with a great deal of operational significance in connection with experiments. In fact, however, contemporary thermodynamics could more appropriately be described as a study of the intrinsic properties of matter. Certainly intrinsic properties of matter should depend neither upon the surrounding partitions nor upon heat engines. To say they are so dependent is akin to saying that the intrinsic properties of matter are dependent upon boundary conditions, and hence these properties are not really very intrinsic. It follows that it should be possible to study some of these intrinsic properties without reference to boundary conditions. This is not to say that partitions cannot be used to deduce useful information about intrinsic properties, for, indeed, they have been used in just this way. The question is, rather, how much information can be deduced without their use. To clarify this outlook, consider an example from the study of the solutions of ordinary, linear, differential equations. One can study the solutions of a particular differential equation in two ways. Either one can attempt to find all the independent solutions and study their properties, or else one can select a particular solution from the infinite set of all solutions by imposing boundary conditions and then study the properties of this solution. Both methods yield useful information about the properties of the solutions of the differential equation. The first procedure, if it can be carried out, is more economical in its use of concepts for it has no need to introduce the idea of boundary conditions. But, by itself, it cannot answer the question of how the solution varies when boundary conditions vary. However, it is easy to see that the information generated by the first method can be applied to a study of boundary value questions by subsequently adding boundary conditions.

Even though the second law of thermodynamics receives much of the blame for the difficulties associated with the understanding of thermodynamics, a portion of the onus must be placed on the first law. The reason is that the first law is one relation connecting three quantities, heat, work, and the internal energy increment and can be used as the definition of only one of them; the remaining two must be given independent definitions. Thermodynamics tries to use the first law to define both the internal energy increment and the heat and attempts to give an independent definition only for work. But there are problems even with the independent definition of work. Clearly, there is no difficulty when one is dealing solely with mechanical work and, of course, the early experiments which served as a foundation for the first law did rely on mechanical work. Subsequently thermodynamics was applied to phenomena involving chemical reactions and electromagnetic fields and in both cases it was necessary to introduce an appropriate expression for work. The definition of "chemical work" was an ad hoc procedure which presupposed the existence of an entropy and which seemed to work quite well. The definition of "electromagnetic work" was, and is, complicated by the existence of several alternative definitions.

Reformulations of thermodynamics have appeared

from time to time; however, none of them specifically deal with the questions I have raised. Can a formulation of thermodynamics be given which is totally devoid of partitions, composite systems, and heat engines? Does such a formulation clarify the concept of entropy? Does it expose any assumptions which are implicitly made but not explicitly stated in the conventional formulations of thermodynamics? Does it provide a sufficiently general expression for thermodynamic work?

Many more objectives could be listed for a theory of thermodynamics in addition to those posed by the foregoing questions. High on the list would be an answer to the question of how dependent is thermodynamics on the form of the expression for work. Another important objective would be the recognition, from the outset, that thermodynamic states of matter are, after all, rather special states. They are special in the sense that they can be completely described by a small number of parameters in contrast with more general states which require a large or perhaps even an infinite number of parameters for their description. The theory should interrelate the two types of states. Certainly somewhere on the list of objectives for a thermodynamic theory would be the removal of the temperature variable from its position of preeminence among other thermodynamic variables. Since this preeminence was achieved largely from the zeroth law, it can be rescinded only by an abrogation of this law. In essence, does a viable thermodynamics exist which includes the notion of temperature but excludes the zeroth law? Finally, a theory should both possess intuitive appeal and assiduously avoid implicit assumptions. This combination of intuitive appeal and explicit assumption enables one to delimit rationally the region of validity of the theory by experimental testing of the assumptions built into the theory.

Only some of these objectives indicate a suitable starting point for a formal theory of thermodynamics. Others merely serve as guideposts in the development of that theory. For example, it is obvious that a theory without partitions and composite systems must effectively be a local theory of thermodynamics. Concomitantly, a local theory has no need for the zeroth law since the zeroth law deals with thermal equilibrium among three systems which are formed pairwise into composite systems. Thermodynamic states can be recognized as special states simply by partitioning the collection of all states into thermodynamic and nonthermodynamic states. By contrast, it is possible to establish the degree of dependence of thermodynamics on the form of the work expression, partitions, and composite systems only by avoiding any statements about them as long as possible. Similarly, to endow a theory with intuitive appeal one should work with concepts that are readily interpretable experimentally and that are likely to be familiar to many people. That is, there should be an almost obvious correspondence between the elements of the mathematical theory and their physical realizations. Furthermore, the collection of physical states should be given a minimal mathematical structure and additional structure should be imposed only as the need arises. This implies that one should

try to establish as much of thermodynamics as possible by purely algebraic arguments since some algebraic structure is certainly a prerequisite for topological structure. It is not at all obvious that the set of thermodynamic states need be given any topology, and certainly it is less obvious just what that topology should be. If topological considerations become a necessity, then a reasonable attempt should be made to justify the chosen topology by making some concrete statements about it. This attitude would prevent the ad hoc assumption of, say, a metrizable topology without justification. It also would prevent the use of an arbitrary topology for the deduction of thermodynamic consequences from topological axioms which might not be satisfied by a particular topology or, indeed, by any topology.

Despite the long history of classical thermodynamics, only a comparatively small number of distinctly different theoretical formulations of thermodynamics exist. These theories will not be discussed in detail, but I shall be content merely to point out a few significant features. Undoubtedly the best known of all thermodynamic theories is the one associated with the names Clausius, Kelvin, and Carnot. It is by far the most common and its exposition can be found in many thermodynamic texts. For example, it can be found in the books by Adkins,[2] Wilson,[3] and Zemansky.[4] This description of thermodynamics is replete with partitions, composite systems, and heat engines. A comparatively more recent theoretical structure is based on the work of Caratheodory[5] more than 60 years ago. More recent expositions have been written by Landsberg,[6] Buchdahl,[7] Bernstein,[8] and Boyling.[9] In some respects this thermodynamics is very similar to the Clausius, Kelvin, Carnot formulation, complete with partitions and composite systems. It differs from the latter chiefly in its approach to the second law, which is obtained not from a heat engine but from an axiom that contains topological considerations. The axiom could aptly be called the adiabatic inaccessibility axiom. The chief function of the axiom is to assure the integrability of a Pfaffian form, the first law of thermodynamics, and so to introduce the entropy. Landsberg assumes the topology to be a metric topology and apparently so do Buchdahl and Bernstein. Boyling (1968) assumes a separable topological space, while Boyling (1972) uses a differentiable manifold and hence a separable Hausdorff topology.

Yet another theory of thermodynamics is the work of Tisza[10] and Callen.[11] It differs significantly from other theories because it presupposes the existence of an entropy with prescribed properties and makes no attempt to derive either the entropy or its properties from more fundamental assumptions. It is an excellent pedagogical device, particularly in Callen's presentation, since it systematizes a great deal of the thermodynamic formalism used in applications. This thermodynamics, like most others, relies heavily on partitions and composite systems.

Numerous conceptual similarities exist between the theory of Falk and Jung[12] and that of Giles,[13] although the concepts are implemented differently. Both theories

place a heavy reliance upon the composite system and both represent processes as ordered pairs of thermodynamic states. Neither introduces entropy through integrability of a Pfaffian form. Significantly, both theories are considerably more algebraic in character than most other theories of thermodynamics, although they are not wholly independent of topological considerations.

A degree of kinship is also exhibited among the thermodynamic theories of Buchdahl,[14] Buchdahl and Greve,[15] Rastall,[16] and Cooper.[17] These authors base their analyses on the adiabatic inaccessibility axiom of Caratheodory. Buchdahl and Buchdahl and Greve use it in connection with a separable, connected, metric space, Rastall employs it in an arbitrary topological space, and Cooper uses it in a separable topological space. In fact, Buchdahl (1962) assumes the usual topology on a Cartesian product of the reals. In each paper the topological structure is involved in the deduction of an entropy; however, this is accomplished without Pfaffian forms in contrast to Caratheodory-like theories. All four papers use adiabatic processes; additionally Rastall introduces so-called energic processes which are generalizations of Buchdahl's[14] isometric processes. Again all four papers utilize composite systems, although Rastall points out that he makes only sparing use of them. Hornix[18] attempts to establish a connection between the thermodynamics represented by the work of Falk and Jung, Giles, and Cooper and the Tisza—Callen approach to thermodynamics. Although Hornix does not specify his topology, his reliance on the work of Cooper implies the use of a separable topology.

My approach to the theory of thermodynamics, presented on the following pages, represents an attempt to deal with the previously stated objectives. It is broadly subdivided into four parts and each portion deals with a different aspect of the problem. Thus, successive sections deal with algebraic, topological, integrability and continuum considerations. Each section is independent of all succeeding sections but draws freely on concepts introduced in preceding sections. In fact, the physical interpretation of a concept becomes progressively more transparent the farther one proceeds with the development of the theory. The reason for this is that only those aspects of a concept are introduced in a given section which are needed in that section; additional facets of the same concept are defined in subsequent sections as they become necessary. The rationale for this fragmentary method of definition is that is permits one to draw conclusions subject to the least restriction and hence capable of the widest applicability.

Instead of immediately proceeding to the development of the theory, let us briefly digress to discuss a related topic. We are considering the formulation of a physical theory and hence it would be well to reflect on the nature of physical theory and attempt to delineate what we can or cannot expect of it. Broadly speaking, a physical theory might be characterized as a mathematical structure capable of mirroring phenomena, thermodynamic phenomena in our case, that occur in

the real world. Thus, in some sense the mathematical structure and the real world "coincide" for a physical theory. But in what sense can the two be the same since it is obvious that mathematics can and does exist independently of the real world and certainly the real world exists without reference to any mathematical structure. The various "morphisms" such as homeomorphisms and several kinds of homomorphisms and isomorphisms provide a precise definition of the equivalence of mathematical structures in mathematics. There is no counterpart which serves the same function for a mathematical notion and a physical phenomenon. In a sense the correspondence between the two is established directly by using names for the mathematical objects which are also interpretable in the real world. Thus, it is possible to conclude that the function of a physical theory is not to give a physical interpretation of a mathematical object but to serve as a predictive device for physical measurements, which are the source of our quantitative knowledge about the real world. As an example, we cannot expect the theory to define equilibrium states unless they are definable and recognizable by measurement. Nor can we expect to give a physical interpretation of a wholly mathematical concept unless that concept can be reduced to the concepts which are interpretable as measurements. It is important to remember that an acceptable physical theory must necessarily be based upon impeccable mathematics; by contrast a mathematical structure may represent perfectly respectable mathematics but not necessarily respectable physical theory.

I. ALGEBRAIC CONSIDERATIONS

Conditioned as we are to expect physical properties to be representable by continuous functions, it might seem strange at first to expect to construct a thermodynamics which is independent of the topological notions of continuity and nearness. Nonetheless, it is possible to cast a substantial portion of thermodynamics into a purely algebraic mold by using only quite elementary algebraic ideas. Most of the prerequisite material is available in any one of a number of introductory texts on abstract algebra; an excellent choice would be the book by Paley and Weichsel.[19] Two options are available in the construction of algebraic thermodynamics. The first is to develop the mathematical and physical ideas simultaneously, while the second is to separate the mathematics and physics by as wide a gulf as possible. I choose to proceed by the second mode since it possesses the advantage of cleanly separating the two kinds of ideas and results in a precise pinpointing of the physical assumptions. Furthermore, the development of the physical ideas is then unencumbered by mathematical asides. The chief disadvantage is the difficulty in supplying an adequate motivation for the reader whose primary concern is the physics and not the mathematics. I feel that the advantages of the chosen approach outweigh the disadvantages, for my aim is to clarify the fundamentals of thermodynamics and such a purpose is not necessarily consonant with pedagogical aims. Nevertheless, some motivation is desirable and, indeed, essential in order to indicate the purpose behind some of the mathematical definitions and theorems

which will precede the thermodynamics. The mathematics to be developed will involve two physically relevant concepts which are of fundamental importance for thermodynamics. One is the notion of states of a physical system and the other is the idea of processes connecting these states. The development of the mathematics of these two concepts will concentrate on two topics. The first is a complete and precise characterization of reversible processes and an elucidation of their function with respect to the states. The second is a similarly careful treatment of the ordering of the states and a determination of the connection between order properties and processes. Once this has been achieved, the thermodynamics can be developed in a relatively straightforward manner.

States and processes will be an integral part of the definition of the mathematical object, called a physical system, which will be the subject of our mathematics. For the present these two notions will be treated as undefined concepts since their detailed definition will in no way affect the algebra. However, we shall become more definite about them as the analysis progresses, and ultimately they will assume the roles conventionally assigned to them.

Axiom I.1: A physical system is the quintet (X, Σ, Π, F, f), where (1) X is a set called the set of all physical states, (2) $\Sigma \subset X$ is a subset of X called the set of all thermodynamic states, (3) Π is a set called the set of all simple processes on Σ, (4) F is a function $F : \Pi \to \Sigma \times \Sigma$, and (5) f is a function $f : \Pi \xrightarrow[\text{onto}]{1-1} \Pi$ satisfying $f \circ f = 1_\Pi$, where 1_Π is the identity map on Π. If p is a simple process, that is, $p \in \Pi$ and $F(p) = (a, b)$, then a and b are said to be the initial and final points, respectively, of p. Further, $f(p) = p^*$ is called the reverse of p, and $F(p) = (a, b)$ if and only if $F(p^*) = (b, a)$.

Henceforth I shall use $\in$ for *is a member of*, $\forall$ in place of *for all*, $\exists$ for *there exists*, $\Rightarrow$ for *implies*, and either iff or $\Longleftrightarrow$ for *if and only if*. The negation of these symbols will be indicated by $/$, the null set will be represented by $\emptyset$, and the logical or will be used in the sense of either A or B or both. Some mental imagery would undoubtedly enhance the understandability of Axiom I.1 as well as the subsequent algebraic manipulations. For this reason, and for no other, it is helpful to view states as points in some finite-dimensional space and simple processes as directed curves connecting pairs of these points, the initial and final points of the process. This is illustrated in Fig. 1. We are now in a position to make some observations about the content of Axiom I.1. First, it presupposes that a distinction already exists between physical states and thermodynamic states, although it does not exclude the possibility that they coincide ($X = \Sigma$). Second, it restricts initial and final points of simple processes to the set Σ (since the range of F is $\Sigma \times \Sigma$) rather than permitting them to be arbitrary physical states. However, in terms of the mental image of a process, it does not require the directed curve to lie within Σ nor does it require p and p^* to be superimposable as shown in Fig. 1. Third, it does not assume that only one simple process can have $a, b \in \Sigma$ as initial and final points

in Σ, nor does it assume that each pair of points in Σ is connected by some simple process; equivalently F has not been assumed to be 1—1 and onto. Such an assumption would contradict experience since there are many routes one can take to convert a gram of ice to a gram of steam, for example. Finally, the adjective simple is appended to the word process merely because I wish to reserve the term process for a combination of simple processes which will be introduced shortly.

Algebra of reversible processes

I shall now begin to develop the mathematical machinery that will be used to erect the algebraic structure of thermodynamics upon the foundation established by Axiom I.1. This subsection and the following one will examine to what extent it is possible to use processes to induce algebraic structure on Σ. Here we will use reversible processes to induce equivalence relations on Σ. In the following subsection processes will be used to induce order properties on Σ. The material in these two subsections will be of a purely mathematical nature and will not utilize any thermodynamic concepts or, indeed, any physical concepts whatsoever apart from those introduced by Axiom I.1. For example, the first two theorems are immediate consequences of Axiom I.1.

Theorem I.2: $(p^*)^* = p$ $\forall$ $p \in \Pi$. Also $p_1 = p_2$ iff $p_1^* = p_2^*$ $\forall$ $p_1, p_2 \in \Pi$.

Proof: $(p^*)^* = f(p^*) = f(f(p)) = f \circ f(p) = 1_\Pi(p) = p$. The remainder of the theorem follows from the fact that f is a 1—1 function. Thus, $p_1 = p_2 \Rightarrow f(p_1) = f(p_2)$ since f is a function, and $f(p_1) = f(p_2) \Rightarrow p_1 = p_2$ since f is 1—1.

Theorem I.3: The map F induces an equivalence relation $\approx$ on Π. That is, $\forall$ $p_1, p_2 \in \Pi$, the relation $\approx$ defined by $p_1 \approx p_2$ iff $F(p_1) = F(p_2)$ is an equivalence relation on Π.

Proof: An equivalence relation is reflexive, symmetric, and transitive. The relation $\approx$ is (1) reflexive since $F(p) = F(p)$ $\forall$ $p \in \Pi$ because F is a function and therefore $p \approx p$. (2) symmetric since $p_1 \approx p_2 \Rightarrow F(p_1)$
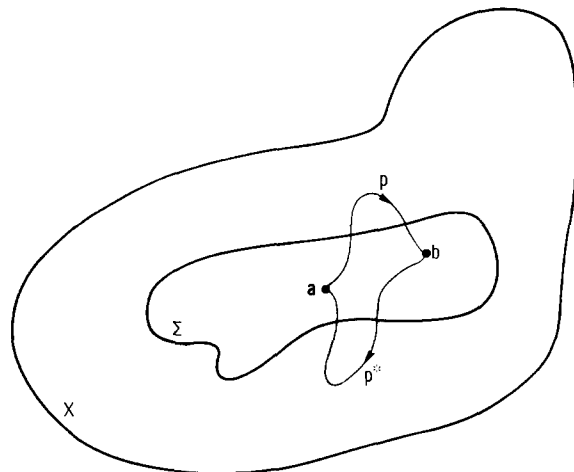
FIG. 1. An example of a simple process p, with initial point a and final point b, and its reverse process p^*.

$=F(p_2) \Rightarrow F(p_2)=F(p_1) \Rightarrow p_2 \approx p_1$, (3) transitive since $p_1 \approx p_2$ and $p_2 \approx p_3 \Rightarrow F(p_1)=F(p_2)$ and $F(p_2)=F(p_3)$ $\Rightarrow F(p_1)=F(p_3) \Rightarrow p_1 \approx p_3$.

As a result of Theorem I.3 it is apparent that the equivalence classes of the relation $\approx$ can be put into a 1−1 correspondence with ImF, the range of F, even though it may not be possible to do this with Π itself.

Definition I.4: A simple process $p \in \Pi$ is said to be null, or a simple null process, iff $p^* = p$.

Theorem I.5: Let $p \in \Pi$. If p is null, then $F(p)=(a,a)$ for some $a \in \Sigma$. If $F(p)=(a,a)$ for some $a \in \Sigma$, then $p^* \approx p$.

Proof: Suppose $p^* = p$. Then $(a,b)=F(p)=F(p^*)$ $=(b,a)$ and therefore $a=b$. From Axiom I.1, $F(p)$ $=(a,b)$ iff $F(p^*)=(b,a)$ and therefore $b=a \Rightarrow F(p)$ $=F(p^*)$ and hence $p \approx p^*$.

Thus, while the initial and final points coincide for all simple null processes, if the initial and final points coincide, then we may only conclude that the simple process and its reverse are equivalent but not necessarily equal.

Subsets of Π play an extremely important role in the algebraic theory being developed. Indeed, it would be difficult to overemphasize their importance. Substantially all of the following definitions and theorems will be statements relative to some subset of Π. The first of these is

Definition I.6: Let $\beta \subset \Pi$ be a collection of simple processes. A simple process $p \in \Pi$ is said to be reversible in β if $p \in \beta$ and $p^* \in \beta$. If p is not reversible in β, then it is said to be irreversible in β.

It is a trivial consequence of the definition of p^* that $\forall p \in \Pi$, p is reversible in Π. This is not, however, true for other collections of simple processes as is shown in the next theorem.

Definition I.7: If $\beta \subset \Pi$, then $\beta^* \subset \Pi$ is defined by $\beta^* = \{p | p \in \Pi, \ p^* \subset \beta\}$.

Theorem I.8: If $\beta \subset \Pi$, then, for any simple processes p, (1) $p^* \in \beta^*$ iff $p \in \beta$, (2) $p^* \in \beta$ iff $p \in \beta^*$, (3) $(\beta^*)^* = \beta$, and (4) p is reversible in β iff $p \in \beta \cap \beta^*$.

Proof: By Definition I.7 we have $p^* \in \beta^*$ iff $p^* \in \Pi$ and $(p^*)^* = p \in \beta$ where Theorem I.2 was used. To establish (3) consider that $(\beta^*)^* = \{p | p \in \Pi, \ p^* \in \beta^*\}$ $= \{p | p \in \Pi, \ p \in \beta\} = \beta$, where part (1) was used together with Definition I.7. To establish (2) we can use (3) and (1) since $p^* \in \beta$ iff $p^* \in (\beta^*)^*$ iff $p \in \beta^*$. The last part of the theorem follows by using part (2) and Definition I.6 since $p \in \beta \cap \beta^*$ iff $p \in \beta$ and $p \in \beta^*$ iff $p \in \beta$ and $p^* \in \beta$, and hence p is reversible in β.

We observe two things. First, it should be obvious from Theorem I.8(4) that p is irreversible in β iff $p \in \beta - \beta \cap \beta^*$, the relative complement of $\beta \cap \beta^*$ in β. Second, the symbol * applied to a subset of Π can be viewed as a map induced by f on the power set of Π, that is, the collection of all subsets of Π. A trivial consequence of Definition I.7 is

Theorem I.9: Let $\mathcal{O} \in \Pi$ be the set of all simple null processes. Then $\mathcal{O}^* = \mathcal{O}$.

Proof: By Definition I.4, $p \in \mathcal{O}$ iff $p = p^*$ and therefore $\mathcal{O}^* = \{p | p \in \Pi, \ p^* = p \in \mathcal{O}\} = \mathcal{O}$.

Further development of the reversible−irreversible concept is possible and, indeed, essential if it is to be utilized in a thermodynamic theory. Perhaps the most obvious extension would be to consider the reversibility of arbitrary subsets of Π rather than the one element subset $\{p\}$. Thus, we make

Definition I.10: Let $\beta \subset \Pi$, let Δ be an index set, let $\{C_\alpha | C_\alpha \subset \Pi, \ \alpha \in \Delta\}$ be the collection of all subsets of Π such that p is reversible in $C_\alpha \ \forall \ p \in \beta$, and let $\{D_\alpha | D_\alpha \subset \beta, \ \alpha \in \Delta\}$ be the collection of all subsets of β such that p is reversible in $\beta \ \forall \ p \in D_\alpha$. The reversible superset of β is the set $(\beta)^r \equiv \cap_{\alpha \in \Delta} C_\alpha$, and the reversible subset of β is the set $(\beta)_r \equiv \cup_{\alpha \in \Delta} D_\alpha$.

Theorem I.11: Let $\beta \subset \Pi$. Then (1) $(\beta)^r$ and $(\beta)_r$ are unique, (2) if $C \subset \Pi$ such that p is reversible in $C \ \forall \ p \in \beta$, then $(\beta)^r \subset C$, (3) if $D \subset \beta \subset \Pi$ such that p is reversible in $\beta \ \forall p \in D$, then $D \subset (\beta)_r$, and (4) $(\beta)^r = \beta \cup \beta^*$ and $(\beta)_r = \beta \cap \beta^*$. Also $(\beta)^r = (\beta)_r$ iff $\beta = \beta^*$.

Proof: To establish (2), observe that if C is a subset of Π such that all elements of β are reversible in C, then $C \in \{C_\alpha | \alpha \in \Delta\}$ and by the property of intersections $(\beta)^r = \cap_{\alpha \in \Delta} C_\alpha \subset C$. Part (3) can be establish by supposing that each element of $D \subset \beta$ is reversible in β. Then $D \in \{D_\alpha | \alpha \in \Delta\}$ and by the property of unions $D \subset \cup_{\alpha \in \Delta} D_\alpha = (\beta)_r$. To prove the uniqueness of $(\beta)^r$, notice that $\forall p \in \beta$, p is reversible in $C_\alpha \ \forall \ \alpha \in \Delta$, and hence p is reversible in $\cap_{\alpha \in \Delta} C_\alpha = (\beta)^r$. If $\bar{B}$ is another reversible superset of β, then by (2) we have $\bar{B} \subset (\beta)^r$ and $(\beta)^r \subset \bar{B}$, and thus $\bar{B} = (\beta)^r$. The uniqueness of $(\beta)_r$ is derived analogously since $\forall p \in (\beta)_r$, $p \in D_\alpha$ for some $\alpha \in \Delta$ and hence p is reversible in β. If $\tilde{B}$ is another reversible subset of β, then from (3) we have $\tilde{B} \subset (\beta)_r$ and $(\beta)_r \subset \tilde{B}$, and therefore $\tilde{B} = (\beta)_r$. Only (4) remains undone. From the definition of β^* it follows that $\beta \cup \beta^* \in \{C_\alpha | \alpha \in \Delta\}$ and by (2) $(\beta)^r \subset \beta \cup \beta^*$. Now $p \in \beta$ is reversible in C_α iff $p, p^* \in C_\alpha$ by Definition I.6. Since this must be true $\forall p \in \beta$, it follows that $\beta \subset C_\alpha$, $\beta^* \subset C_\alpha \ \forall \ \alpha \in \Delta$, and therefore $\beta \cup \beta^* \subset C_\alpha$ $\forall \ \alpha \in \Delta$ and $\beta \cup \beta^* \subset \cap_{\alpha \in \Delta} C_\alpha = (\beta)^r$. Thus $(\beta)^r = \beta \cup \beta^*$. From Theorem I.8(4) it follows that $\beta \cap \beta^* = \{D_\alpha | \alpha \in \Delta\}$ and by (3) $\beta \cap \beta^* \subset (\beta)_r$. Suppose $p \in (\beta)_r$; then by definition of $(\beta)_r$ it follows that p is reversible in β and by Theorem I.8(4) $p \in \beta \cap \beta^*$, and therefore $(\beta)_r \subset \beta \cap \beta^*$, and then $(\beta)_r = \beta \cap \beta^*$. Finally, suppose $\beta = \beta^*$; then $\beta \cup \beta^* = \beta \cap \beta^*$. Conversely suppose $\beta \cup \beta^* = \beta \cap \beta^*$ and $\beta \neq \beta^*$. Then $\beta \cup \beta^* - \beta \cap \beta^* = \emptyset$ and $\exists \ p \in \beta$ and $p \notin \beta^*$ and therefore $p \in \beta \cup \beta^*$ and $p \notin \beta \cap \beta^*$, and hence $p \in \beta \cup \beta^* - \beta \cap \beta^*$, which contradicts $\beta \cup \beta^* - \beta \cap \beta^* = \emptyset$.

This theorem has now established that for any subset $\beta \subset \Pi$, there exists a largest subset $(\beta)_r$ such that all elements of $(\beta)_r$ are reversible in β and, further, there exists a smallest superset $(\beta)^r$ with the property that all members of β are reversible in $(\beta)^r$. What about the reversible subsets and supersets of $(\beta)^r$ and $(\beta)_r$? This is answered by the following corollary.
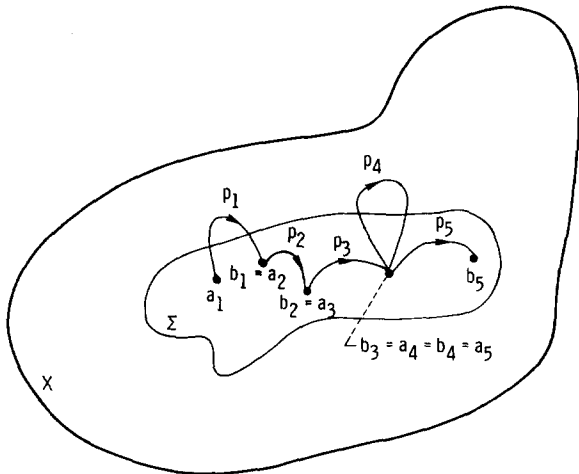
FIG. 2. A process on Σ of length five with initial point a_1 and final point b_5. For clarity the simple null processes after p_5 are not shown.

Corollary I. 12: Let $\beta \subset \Pi$. Then $[(\beta)^r]_r = (\beta)^r = [(\beta)^r]^r$ and $[(\beta)_r]_r = (\beta)_r = [(\beta)_r]^r$.

Proof: If $\beta = \beta^*$, then $\beta \cup \beta^* = \beta$ and $\beta \cap \beta^* = \beta$. Hence, if it can be shown that $[(\beta)^r]^* = (\beta)^r$ and $[(\beta)_r]^* = (\beta)_r$, then this corollary is a direct consequence of Theorem I. 11(4). The definition of the * operation applied to sets is given in Definition I. 7 and, applying it to $(\beta)^r$, we find $[(\beta)^r]^* = \{p \mid p \in \Pi, \ p^* \in (\beta)^r = \beta \cup \beta^*\}$. But using Theorem I. 8(1, 2), we obtain $p^* \in \beta \cup \beta^*$ iff $p^* \in \beta^*$ or $p^* \in \beta$ iff $p \in \beta$ or $p \in \beta^*$ iff $p \in \beta \cup \beta^*$. Hence, $[(\beta)^r]^* = \{p \mid p \in \Pi, \ p \in \beta \cup \beta^* = (\beta)^r\} = (\beta)^r$. In a like manner $[(\beta)_r]^* = \{p \mid p \in \Pi, \ p^* = (\beta)_r = \beta \cap \beta^*\}$. But $p^* \in \beta \cap \beta^*$ iff $p^* \in \beta$ and $p^* \in \beta^*$ iff $p \in \beta^*$ and $p \in \beta$ iff $p \in \beta \cap \beta^* = (\beta)_r$, and thus $[(\beta)_r]^* = (\beta)_r$.

Ultimately it will become necessary for us to have available a generalization of the notion of a simple process. This generalization is supplied by the next definition.

Definition I. 13: Let N be the set of positive integers. A sequence of simple processes is a map $P: N \to \Pi$, where $P(n) = p_n \in \Pi$. A sequence of simple processes is said to be finite of length $n_0 \in N$ iff p_{n_0} is not null and p_n is null $\forall \ n > n_0$. It is said to be finite of length zero iff p_n is null $\forall \ n \in N$. A process P on Σ is a finite sequence of simple processes such that $b_n = a_{n+1} \ \forall \ n \in N$, where $F(p_n) = (a_n, b_n)$. More particularly, a process P such that $\text{Im} P \subset \beta \subset \Pi$ is called a β-process on Σ. If P is a β-process on Σ of length $n_0 \in N$, then P is said to β-link $x, x' \in \Sigma$ iff $x = a_1$ and $x' = b_{n_0}$. This is represented by $x \xrightarrow{P} x'$, and x, x' are the initial and final points of process P.

An example of a process of length five is schematically shown in Fig. 2, where, for clarity, all simple null processes after p_5 have been excluded from the sketch. However, note that from Theorem 1.5 p_4 might be a simple null process. It is apparent that each process P induces a sequence in Σ, namely, $g: N \to \Sigma$, where $g(n) = a_n$, but this then raises the possibility of considering a process to be an infinite sequence of simple processes with convergence of this sequence determined by convergence of the induced sequence in

Σ. While this is certainly possible, such a procedure would violate my avowed intention to eschew topology in Σ; convergence of an infinite sequence in Σ can only be defined in terms of topological ideas in Σ. Finally, the fact that each process P has an initial and final point can be regarded as an extension of the map F to the set of all processes from the set of all processes of length one since a process of length one is effectively a simple process. Equivalently, F induces a map from the set of all processes to $\Sigma \times \Sigma$.

Naturally the observation that F can be extended from simple processes motivates us to look for other extensions. The extension of the concepts of null processes and reversible processes is accomplished in

Definition I. 14: Let P be a process of length n_0 on Σ. The process P is said to be (1) null iff $n_0 = 0$ and (2) reversible in $\beta \subset \Pi$ iff $P(n)$, $P^*(n) \in \beta \ \forall \ n \in N$. The process P is irreversible in β iff it is not reversible in β. The reverse of P is a process $P^*: N \to \Pi$, where $P^*(n) = p^*_{n_0+1-n}$ for $n \leq n_0$ and $P^*(n)$ is null for $n > n_0$.

Note that for any process P there might correspond many reverse processes since no restriction was placed on the choice of null processes $P^*(n)$ for $n > n_0$. With this in mind it is possible to derive a theorem analogous to Theorem I. 8.

Theorem I. 15: If $\mathcal{O} \subset \Pi$ is the set of all simple null processes and $\mathcal{O} \subset \beta \subset \Pi$, then for any process P (1) $\text{Im} P^* \subset \beta^*$ iff $\text{Im} P \subset \beta$, (2) $\text{Im} P^* \subset \beta$ iff $\text{Im} P \subset \beta^*$, and (3) P is reversible in β iff $\text{Im} P \subset \beta \cap \beta^* = (\beta)_r$.

Proof: As was pointed out above, $P \ (n > n_0) \in \mathcal{O} \cap \beta$ does not imply $P^* \ (n > n_0) \in \mathcal{O} \cap \beta^*$. However, if $\mathcal{O} \subset \beta$, then $\mathcal{O} = \mathcal{O}^* \subset \beta^*$ and, conversely, if $\mathcal{O} \subset \beta^*$, then $\mathcal{O} = \mathcal{O}^* \subset \beta^*$ and then $\mathcal{O} \subset \beta$. Thus we have $\mathcal{O} \subset \beta$ iff $\mathcal{O} \subset \beta^*$. However, under these circumstances $\beta \cap \mathcal{O} = \mathcal{O} = \beta^* \cap \mathcal{O}$ and hence $P \ (n > n_0) \in \mathcal{O} \cap \beta$ iff $P^* \ (n > n_0) \in \mathcal{O} \cap \beta^*$. The first two parts of this theorem now follow from Theorem I. 8(1, 2). Now if we use part two of this theorem, we have P is reversible in β iff $P(n)$, $P^*(n) \in \beta \ \forall \ n \in N$ iff $\text{Im} P \subset \beta$ and $\text{Im} P^* \subset \beta$ iff $\text{Im} P \subset \beta$ and $\text{Im} P \subset \beta^*$ iff $\text{Im} P \subset \beta \cap \beta^* = (\beta)_r$.

The following corollary to Theorem I. 15 is of considerable importance for thermodynamics.

Corollary I. 16: If $\mathcal{O} \subset \Pi$ is the set of all simple null processes and $\mathcal{O} \subset \beta \subset \Pi$, then all $(\beta \cap \beta^*)$-processes $[(\beta \cup \beta^*)$-processes$]$ are reversible in $\beta \cap \beta^*$ $[\beta \cup \beta^*]$.

Proof: P is a $(\beta \cap \beta^*)$-process iff $\text{Im} P \subset \beta \cap \beta^* = (\beta)_r$. But from Theorem I. 15(3) we have that P is reversible in $(\beta)_r$ iff $\text{Im} P \subset [(\beta)_r]_r = (\beta)_r$, where Corollary I. 12 was used in the last step. Hence, all $(\beta \cap \beta^*)$-processes are reversible in $\beta \cap \beta^*$. A similar proof holds for $(\beta \cup \beta^*)$-processes.

The significance of this result lies in the fact that if P is a $(\beta \cap \beta^*)$-process $[(\beta \cup \beta^*)$-process$]$ and $x \xrightarrow{P} x'$, then P^* is also a $(\beta \cap \beta^*)$-process $[(\beta \cup \beta^*)$-process$]$ and $x' \xrightarrow{P^*} x$.

Once we have available Corollary I. 16 it becomes a relatively simple matter to make a connection between processes on Σ and equivalence relations on Σ.

This is demonstrated in the next definition and theorem.

Definition I.17: Let $\beta \subset \Pi$ be a collection of simple processes containing the set O of all simple null processes. A point $x \in \Sigma$ is said to be β-equivalent to $x' \in \Sigma$, $x \stackrel{\beta}{=} x'$, iff $\exists$ a $(\beta \cup \beta^*)$-process P such that $x \stackrel{P}{=} x'$. It is said to be β-equal to x', $x =_\beta x'$, iff $\exists$ a $(\beta \cap \beta^*)$-process P such that $x \stackrel{P}{=} x'$.

Theorem I.18: Let $O \subset \Pi$ be the collection of all simple null processes and suppose that, for each $x \in \Sigma$, $\exists$ some $p \in O$ such that $F(p) = (x, x)$ and $O \subset \beta \subset \Pi$. Then (1) the relations $\stackrel{\beta}{=}$ and $=_\beta$ are equivalence relations on Σ, (2) $x =_\beta x' \Rightarrow x \stackrel{\beta}{=} x'$, that is, each equivalence class of $=_\beta$ is a subset of some equivalence class of $\stackrel{\beta}{=}$, and(3) if $\beta = \beta^*$, then $x =_\beta x'$ iff $x \stackrel{\beta}{=} x'$, that is, if $\beta = \beta^*$, then the equivalence classes of $=_\beta$ coincide with those of $\stackrel{\beta}{=}$.

Proof: We first observe that $O \subset \beta \cap \beta^* \subset \beta \cup \beta^*$. The relation $=_\beta$ is (1) reflexive since for each $x \in \Sigma$ $\exists$ some $p \in O$ such that $F(p) = (x, x)$ and hence let P be the constant function $P : N \to \Pi$ defined by $P(n) = p$ $\forall n \in N$. Then $x \stackrel{P}{=} x$ and $x =_\beta x$. The relation is (2) symmetric since $x =_\beta x'$ iff $\exists$ a $(\beta \cap \beta^*)$-process P such that $x \stackrel{P}{=} x'$, and then, by Corollary I.16, P^* is a $(\beta \cap \beta^*)$-process such that $x' \stackrel{P^*}{=} x$ iff $x' =_\beta x$. The relation is (3) transitive since $x =_\beta x'$ and $x' =_\beta x''$ iff $\exists$ $(\beta \cap \beta^*)$-processes P and P' such that $x \stackrel{P}{=} x'$ and $x' \stackrel{P'}{=} x''$. If P is a process of length n_0, then the process $P'' : N \to \Pi$ defined by $P''(n) = P(n)$ for $n \leq n_0$ and $P''(n) = P'(n - n_0)$ for $n > n_0$ is a $(\beta \cap \beta^*)$-process and $x \stackrel{P''}{=} x''$ iff $x =_\beta x''$. The proof that $\stackrel{\beta}{=}$ is an equivalence relation follows in an identical manner. To establish (2), we notice that $\beta \cap \beta^* \subset \beta \cup \beta^*$ implies that each $(\beta \cap \beta^*)$-process is also a $(\beta \cup \beta^*)$-process. Part (3) is true since $\beta = \beta^* \Rightarrow \beta \cap \beta^* = \beta \cup \beta^*$ and hence $x =_\beta x'$ iff $x \stackrel{\beta}{=} x'$.

Theorem I.18 is the culmination of the efforts expended to produce a complete analysis of the notion of a reversible process. It will play a significant role in the development of the algebraic theory of thermodynamics. In connection with this theorem it should be noted that $\beta = \beta^*$ was sufficient to demonstrate $x =_\beta x'$ iff $x \stackrel{\beta}{=} x'$. However, asserting $\beta \neq \beta^*$ does not assure us that the equivalence classes of $=_\beta$ and $\stackrel{\beta}{=}$ do not coincide, because it certainly is possible that for each $(\beta \cup \beta^*)$-process there exists some $(\beta \cap \beta^*)$-process linking the same points while still maintaining $\beta \neq \beta^*$.

The algebra of partial order relations

Recall that in addition to the algebra of reversible processes we must still make a careful analysis of order properties before we can attempt to get into the details of thermodynamics. The treatment of order properties to be given here represents only a modest generalization of the discussion found in textbooks. This generalization, however, is essential for the application to thermodynamics and, indeed, for the concept of entropy used in this paper. Order properties can be phrased in terms of partial orders and strict orders. The former is defined by

Definition I.19: A partially ordered set is a triplet $(M, \leq_M, =_M)$, where M is a set, $=_M$ is an equivalence relation on M, and $\leq_M$ is a relation on M which satis-

fies (1) $m \leq_M m$ $\forall$ $m \in M$ (reflexive), (2) $m \leq_M m'$ and $m' \leq_M m'' \Rightarrow m \leq_M m''$ (transitive), and (3) $m \leq_M m'$ and $m' \leq_M m \Rightarrow m =_M m'$ (antisymmetric). The relation $\leq_M$ is called a partial order on M. Elements m, $m' \in M$ are said to be comparable iff $m \leq_M m'$ or $m' \leq_M m$. M is said to be a chain, or M is said to be linearly ordered by $\leq_M$ iff $\forall$ m, $m' \in M$, $m \leq_M m'$ or $m' \leq_M m$. If M is linearly ordered by $\leq_M$, then $\leq_M$ is called a linear (or total) order relation on M.

In the conventional definition of partial order the equivalence relation of ordinary equality is used in place of the arbitrary equivalence relation $=_M$ that was used in Definition I.19. This is the only difference between Definition I.19 and the usual one and is the generalization referred to previously. Now any subset of a partially ordered set M can itself be regarded as a partially ordered set, ordered by the restriction of $\leq_M$ and $=_M$ to the subset. Thus, a subset inherits a partial order. With this convention a subset of a partially ordered set M might be a chain even though M itself is not. Of course, every subset of a chain is obviously a chain. To complement the definition of partial order we have the conventional definition of strict order.

Definition I.20: A strictly ordered set is a pair $(M, <_M)$, where M is a set and $<_M$ is a relation on M which satisfies (1) $m <_M m'$ and $m' <_M m'' \Rightarrow m <_M m''$ (transitive) and (2) $m <_M m' \Rightarrow m' \not<_M m$. The relation $<_M$ is called a strict order on M and M is said to be strictly ordered.

Manifest differences exist between the definition of partial order and that of strict order. These differences, however, are more apparent than real, for partial order can be used to induce strict order and conversely. This duality is established by the following two theorems.

Theorem I.21: Let $(M, \leq_M, =_M)$ be a partially ordered set and define a relation $<_M$ by $m <_M m'$ iff $m \leq_M m'$ and $m \neq_M m'$. The relation $<_M$ is the strict order induced by $\leq_M$ and $=_M$ if (1) $m =_M m' \Rightarrow m \leq_M m'$ and $m' \leq_M m$ or (2) $(m \leq_M m'$ and $m' =_M m'')$ or $(m =_M m'$ and $m' \leq_M m'') \Rightarrow m \leq_M m''$.

Proof: Let $m <_M m'$ and, by the definition of $<_M$, this $\Rightarrow m \leq_M m'$ and $m \neq_M m'$. Suppose $m' <_M m$. Hence $m' \leq_M m$ and $m' \neq_M m$. But $m \leq_M m'$ and $m' \leq_M m \Rightarrow m =_M m'$, a contradiction. Thus, $m <_M m' \Rightarrow m' \not<_M m$. To establish the transitivity of $<_M$, suppose $(m <_M m'$ and $m' <_M m'') \Rightarrow (m \leq_M m'$, $m \neq_M m'$ and $m' \leq_M m''$, $m' \neq_M m'') \Rightarrow m \leq_M m''$ by transitivity of $\leq_M$. Suppose $m =_M m''$. By condition (1), $m'' \leq_M m$ and, since $m' \leq_M m''$, $\Rightarrow m' \leq_M m$ and, since $m \leq_M m'$, $\Rightarrow m =_M m'$, which contradicts $m <_M m'$. Hence, $m \leq_M m''$ and $m \neq_M m'' \Rightarrow m <_M m''$, so that $<_M$ is transitive. If condition (2) holds, then $m =_M m''$ and $m \leq_M m' \Rightarrow m'' \leq_M m'$ and, since $m' \leq_M m''$, $\Rightarrow m' =_M m''$ which contradicts $m' <_M m''$. Similarly, $m =_M m''$ and $m' \leq_M m'' \Rightarrow m =_M m'$, which contradicts $m <_M m'$. Again $<_M$ is transitive.

Theorem I.22: Let $(M, <_M)$ be a strictly ordered set and $=_M$ an equivalence relation on M with the properties (1) $m <_M m'$ and $m' =_M m'' \Rightarrow m <_M m''$, (2) $m =_M m'$

and $m' <_M m'' \Rightarrow m <_M m''$ and define a relation $\leq_M$ by $m \leq_M m'$ iff $m <_M m'$ or $m =_M m'$. The relation $\leq_M$ is the partial order induced by $<_M$ and $=_M$.

Proof: The relation $\leq_M$ is (1) reflexive for $m \leq_M m$ because $m =_M m$, (2) antisymmetric for, if $m \leq_M m' \Rightarrow m <_M m'$ or $m =_M m'$ and $m' \leq_M m \Rightarrow m' <_M m$ or $m =_M m'$, then either $m <_M m'$ or $m' <_M m$ but not both because $<_M$ is a strict order. In either event $m =_M m'$. It is (3) transitive for if $m \leq_M m'$ and $m' \leq_M m''$, then $(m <_M m'$ or $m =_M m')$ and $(m' <_M m''$ or $m' =_M m'')$. The four possibilities lead to $m <_M m''$ or $m =_M m'' \Rightarrow m \leq_M m''$.

Examples of the duality between partial order and strict order are easily found, especially when condition (1) of Theorem I.21 is satisifed. In this case it follows that $m \leq_M m'$ and $m' \leq_M m$ iff $m =_M m'$. One such example is supplied by the reals R with the usual order. Here the equivalence relation is the ordinary equality $=$, $r_1 \leq r_2$ iff $r_2 - r_1$ is nonnegative, $r_1 < r_2$ iff $r_2 - r_1$ is positive, and the partially ordered set $(R, \leq, =)$ is a chain. A second example is the power set of any set. The power set is ordered by set inclusion, where $A \subset B$ and $B \subset A$ is used to define $A = B$. Here the partially ordered set is not a chain. Because of the many possibilities for partially ordered sets it becomes important to have some criterion for judging when two partially ordered sets can be regarded as having the same order structure. This is supplied by

Definition I.23: Let g be a function from the partially ordered set $(A, \leq_A, =_A)$ to the partially ordered set $(B, \leq_B, =_B)$ such that (1) $g(a_1) \leq_B g(a_2)$ iff $a_1 \leq_A a_2$ and (2) $g(a_1) =_B g(a_2)$ iff $a_1 =_A a_2$ ∀ $a_1, a_2 \in A$. Then g is called an order homomorphism from A to B, and Img is called the order homomorphic image of A under g. If g is $1-1$, then g is said to be an order isomorphism or a similarity. If Img $\neq B$, then A is said to be order homomorphically (order isomorphically) embedded in B if g is not (is) $1-1$.

Note that if partially ordered sets A and B satisfy Theorem I.21(1), then $a_1 =_A a_2$ iff $a_1 \leq_A a_2$ and $a_2 \leq_A a_1$ iff $g(a_1) \leq_B g(a_2)$ and $g(a_2) \leq_B g(a_1)$ iff $g(a_1) =_B g(a_2)$ and the second condition of Definition I.23 is redundant.

A prime concern in the application of order properties to thermodynamics is the concept of a chain in a partially ordered set. Some standard terminology is needed to facilitate a discussion of some important properties of chains.

Definition I.24: Let $(M, \leq_M, =_M)$ be a partially ordered set. An element $m^0 \in M$ is said to be the largest element of m iff $m \leq_M m^0$ ∀ $m \in M$. It is said to be maximal iff $m^0 \leq_M m \Rightarrow m =_M m^0$. An element $m_0 \in M$ is said to be the smallest element of M iff $m_0 \leq_M m$ ∀ $m \in M$. It is said to be minimal iff $m \leq_M m_0 \Rightarrow m =_M m_0$. Let A be a subset of M. An element $a^0 \in M$ is an upper bound of A iff $a \leq_M a^0$ ∀ $a \in A$. An element $a_0 \in M$ is a lower bound of A iff $a_0 \leq_M a$ ∀ $a \in A$. When they exist, the smallest element of the set of upper bounds of A is denoted by sup(A) while the largest element of the set of lower bounds is denoted by inf(A).

In addition to Definition I.24 we shall also need a well-known tool of mathematics known as Zorn's lemma.

Lemma I.25 (Zorn): Let $(M, \leq_M, =_M)$ be a nonempty partially ordered set in which every chain has an upper bound. Then M contains a maximal element.

Both Definition I.24 and Zorn's lemma are used in the proof of

Theorem I.26: Let $(M, \leq_M, =_M)$ be a partially ordered set. Then (1) $\emptyset$ is the unique minimal chain in M, (2) maximal chains exist in M, and (3) $\emptyset$ is not a maximal chain.

Proof: Let C be the collection of all chains in M, that is, $C = \{C \mid C \subset M, C$ a chain$\}$ and order C by set inclusion. Then $(C, \subset, =)$ is a partially ordered set. Now $\emptyset \in C$ since $\emptyset$ is the null subset of M and has no elements and thus all of its elements are comparable. Further, $\emptyset$ is minimal in C since if $C \in C$ and $C \subset \emptyset$ then, because $\emptyset \subset C$, it follows that $C = \emptyset$. Suppose $\overline{C} \in C$ and $\overline{C}$ is minimal. Then for any chain C such that $C \subset \overline{C}$ we have $C = \overline{C}$. Set $C = \emptyset$ to find $\overline{C} = \emptyset$. To establish the existence of maximal chains in M, let J be a chain in C and let $D = \cup_{C \in J} C$. Suppose $a, b \in D$. Then $a \in C_1$ and $b \in C_2$ for some $C_1, C_2 \in J$ and, since J is a chain, $C_1 \subset C_2$ or $C_2 \subset C_1$ and thus $a, b \in C_2$ or $a, b \in C_1$. But C_1 and C_2 are chains and in either case a and b are comparable. Thus D is a chain in M. Clearly $C \subset D$ ∀ $C \in J$ and D is an upper bound for J. But then every chain in C has an upper bound and, by Zorn's lemma, C possesses a maximal element and thus M has at least one maximal chain. It is obvious that $\emptyset$ is not a maximal chain since for every chain C it is true that $\emptyset \subset C$ and hence this cannot imply $C = \emptyset$ unless all chains in M are null, that is, $M = \emptyset$.

Key results, for thermodynamic purposes, about the properties of partially ordered sets are now almost within reach. In fact, only one more definition must be supplied.

Definition I.27: Let $(M, \leq_M, =_M)$ be a partially ordered set. The relation $\leq_M$ is said to be a nonbranching partial order and the triplet $(M, \leq_M, =_M)$ is said to be a non-branching partially ordered set iff (1) $m_1 \leq_M m_2$ and $m_1 \leq_M m_3 \Rightarrow m_2 \leq_M m_3$ or $m_3 \leq_M m_2$ and (2) $m_2 \leq_M m_1$ and $m_3 \leq_M m_1 \Rightarrow m_2 \leq_M m_3$ or $m_3 \leq_M m_2$ ∀ $m_1, m_2, m_3 \in M$.

A condition similar to Definition I.27 was first used by Falk and Jung[12] (see p. 124 of their paper) and the term nonbranching was used in their paper. Giles[13] (p. 27, Axiom iv) uses the equivalent of part (1) but not part (2). In a sense, Boyling[9] (1968, condition A), (1972, postulate IIa), Buchdahl[14] (1958, Eq. 4.1), (1962, Eq. 4), Buchdahl and Greve[15] (Eq. 3.2), and Rastall[16] (Axiom III) used an assumption that corresponds to a special case of Definition I.27. This will be obvious following Theorem I.30 as will the fact that the term nonbranching is really quite descriptive for a partial order which satisfies Definition I.27. As we pursue the consequences of Definition I.27, it should be kept in mind that $\leq_M$ need not be a total order and hence some points of M may not be comparable.

Theorem I.28: Let $(M, \leq_M, =_M)$ be a nonbranching, partially ordered set. Then (1) if C_1, C_2 are chains in M such that $C_1 \cap C_2 \neq \emptyset$, then $C_1 \cup C_2$ is a chain in M and (2) if C_1, C_2 are maximal chains in M, then $C_1 \cap C_2 = \emptyset$ or $C_1 = C_2$.

Proof: If $a \in C_1 \cap C_2$, then $a \in C_1$ and $a \in C_2$. But
$a \in C_1 \Rightarrow a \leq_M c_1$ or $c_1 \leq_M a$ $\forall$ $c_1 \in C_1$, and $a \in C_2 \Rightarrow a \leq_M c_2$
or $c_2 \leq_M a$ $\forall$ $c_2 \in C_2$. Since $\leq_M$ is nonbranching, the four
possibilities imply that $c_1 \leq_M c_2$ or $c_2 \leq_M c_1$ $\forall$ $c_1 \in C_1$,
$c_2 \in C_2$ and hence $C_1 \cup C_2$ is a chain. To prove (2) let
C_1 and C_2 be maximal chains and suppose $C_1 \cap C_2 \neq \emptyset$.
Then by (1) $C_1 \cup C_2$ is a chain and clearly $C_1 \subset C_1 \cup C_2$
and $C_2 \subset C_1 \cup C_2$. But since C_1 and C_2 are maximal
chains, we have $C_1 \cup C_2 = C_1$ and $C_1 \cup C_2 = C_2$, and
therefore $C_1 = C_2$.

Theorem I.29: If $(M, \leq_M, =_M)$ is a nonbranching
partially ordered set and $C \neq \emptyset$ a chain in M, then $\exists$ a
maximal chain in M containing C.

Proof: Let $\mathcal{J}(C)$ be the collection of all chains in M
that contain C, that is, $\mathcal{J}(C) = \{C_\alpha \subset M \mid C_\alpha$ a chain,
$C \subset C_\alpha$ $\forall$ $\alpha \in \Delta\}$. Obviously $C \in \mathcal{J}(C)$. Let $D = \cup_{\alpha \in \Delta} C_\alpha$
and suppose $a, b \in D$. Then for some $\alpha, \beta \in \Delta$ we have
$a \in C_\alpha$, $b \in C_\beta$. Since $C \subset C_\alpha$ and $C \subset C_\beta$, then $C_\alpha \cap C_\beta \neq \emptyset$
and, by Theorem I.28, $C_\alpha \cup C_\beta$ is a chain, $a, b \in C_\alpha \cup C_\beta$,
and therefore $a \leq_M b$ or $b \leq_M a$. Thus, we see that D is
a chain. To show D is maximal suppose $\overline{C}$ is a chain and
suppose $D \subset \overline{C}$. Then $\overline{C} \cap D \neq \emptyset$ and $\overline{C} \cap D = \overline{C} \cap (\cup_{\alpha \subseteq \Delta} C_\alpha)$
$= \cup_{\alpha \subseteq \Delta}(\overline{C} \cap C_\alpha)$. Hence, for some $\alpha \in \Delta$, $\overline{C} \cap C_\alpha \neq \emptyset$ and,
by Theorem I.27, $\overline{C} \cup C_\alpha$ is a chain containing C since
$C \subset C_\alpha$. This implies that $\overline{C} \cup C_\alpha \in \mathcal{J}(C)$ and then
$\overline{C} \subset \overline{C} \cup C_\alpha \subset D$. But then $\overline{C} \subset D$ and $D \subset \overline{C}$ and, of course,
$\overline{C} = D$. Thus, D is a maximal chain containing C

Theorem I.30: Let $(M, \leq_M, =_M)$ be a partially ordered
set. Then the collection of maximal chains of M is a
partition of M iff $\leq_M$ is nonbranching.

Proof: Let $\mathcal{C} = \{C_\alpha \mid \alpha \in \Delta\}$ be the collection of maximal
chains in M. By Theorem I.26(3), $C_\alpha \neq \phi$ $\forall \alpha \in \Delta$. Since
each $m \in M$ is a member of the one element chain $\{m\}$,
it follows from Theorem I.29 that $m \in C_\alpha$ for some
$\alpha \in \Delta$ and therefore $\cup_{\alpha \in \Delta} C_\alpha = M$. Finally, by Theorem
I.28, if $C_\alpha \neq C_\beta$ then $C_\alpha \cap C_\beta = \emptyset$. But these three condi-
tions are precisely the requirements that make $\mathcal{C}$ a
partition of M. Conversely, suppose $\mathcal{C}$ is a partition
and $m_1, m_2, m_3 \in M$. If $m_1 \leq_M m_2$ and $m_1 \leq_M m_3$, then
$m_1, m_2, m_3 \in C_\alpha$ for some $\alpha \in \Delta$ since elements in dif-
ferent chains are not comparable. Similarly, $m_2 \leq_M m_1$
and $m_3 \leq_M m_1 \Rightarrow m_1, m_2, m_3 \in C_\alpha$ for some $\alpha \in \Delta$. Hence
$m_2 \leq_M m_3$ or $m_3 \leq_M m_2$ and thus $\leq_M$ is nonbranching.

The use of the terminology "nonbranching" to describe
the partial order of Definition I.27 can now be easily
understood. It merely means that comparable elements
can never belong to more than one maximal chain or,
equivalently, that distinct maximal chains never inter-
sect. The assumptions of Boyling,[9] Buchdahl[14] (1958,
1962), Buchdahl and Greve,[15] and Rastall[16] correspond
to the simplest possible case of a nonbranching partial
order, that is, the case where M is the only maximal
chain.

Essentially all mathematical prerequisites for
thermodynamics have now been established. To close
the purely mathematical portion of the paper, I shall
show how processes can be used to induce partial order
relations. This will complement the earlier develop-
ments which related processes to the equivalence
relations $\overset{\beta}{\sim}$ and $=_\beta$.

Definition I.31: Let $\beta \subset \Pi$ be a collection of simple
processes containing the set $\mathcal{O}$ of all simple null pro-
cesses. A point $x \in \Sigma$ is said to be β-comparable to
$x' \in \Sigma$, $x \leq_\beta x'$, iff $\exists$ a β-process P such that $x \overset{P}{\to} x'$.

Theorem I.32: Let $\mathcal{O} \subset \beta \subset \Pi$ and suppose (1) for each
$x \in \Sigma$ $\exists$ some $p \in \mathcal{O}$ such that $F(p) = (x, x)$ and (2) if P is
a β-process such that $x \overset{P}{\to} x'$ and $\overline{P}$ is a β-process such
that $x' \overset{\overline{P}}{\to} x$, then $\exists$ a $(\beta \cap \beta^*)$-process $\tilde{P}$ such that
$x \overset{\tilde{P}}{\to} x'$. Then $(\Sigma, \leq_\beta, =_\beta)$ is a partially ordered set, and
$x =_\beta x'$ iff $x \leq_\beta x'$ and $x' \leq_\beta x$ $\forall$ $x, x' \in \Sigma$. If $\beta = \beta^*$,
then $x \leq_\beta x'$ iff $x =_\beta x'$.

Proof: By Theorem I.18(1) we know that $=_\beta$ is an
equivalence relation and we need only consider the
relation $\leq_\beta$ and show that it satisfies Definition I.19.
Since $\mathcal{O} \subset \beta$, the reflexivity of $\leq_\beta$ follows exactly as
the reflexivity of $=_\beta$ in the proof of Theorem I.18.
Further, the transitivity of $\leq_\beta$ is proven in the same
manner as the transitivity of $=_\beta$. Thus, there only
remains the antisymmetry property. Suppose $x \leq_\beta x'$
$\Longleftrightarrow x \overset{P}{\to} x'$, P a β-process, and $x' \leq_\beta x \Longleftrightarrow x' \overset{\overline{P}}{\to} x$,
$\overline{P}$ a β-process. But then $\exists$ a $(\beta \cap \beta^*)$-process linking
x and x', and hence $x =_\beta x'$. Now if $x =_\beta x'$, then
$\exists$ a $(\beta \cap \beta^*)$-process P such that $x \overset{P}{\to} x'$ and, since
$\beta \cap \beta^* \subset \beta$, $x \leq_\beta x'$. But, by Corollary I.16, P^* is a
$(\beta \cap \beta^*)$-process and $x' \overset{P^*}{\to} x$, and hence $x' \leq_\beta x$. Sup-
pose $\beta = \beta^*$; then $\beta \cap \beta^* = \beta$. Thus, P is a β-process
iff it is a $(\beta \cap \beta^*)$-process and hence $x \leq_\beta x'$ iff $x =_\beta x'$.

The situation that exists in the last part of this theo-
rem is analogous to that in Theorem I.18(3). That is,
$\beta = \beta^*$ is sufficient to demonstrate that there do not
exist elements x, $x' \in \Sigma$ such that $x <_\beta x'$. However,
asserting $\beta \neq \beta^*$ does not assure us that such elements
exist because it is certainly possible that for each
β-process there exists some $(\beta \cap \beta^*)$-process linking
the same points and still maintaining $\beta \neq \beta^*$. Thus,
$\beta \neq \beta^*$ is necessary but not sufficient for the existence
of strictly ordered points.

Algebraic thermodynamics

Let us now proceed to the actual construction of
thermodynamics by implementing the information
reposited in the preceding definitions and theorems.
This can be accomplished by augmenting the assump-
tions represented by Axiom I.1 with additional assump-
tions of a phyiscal nature. Before setting down the
next axiom, let me point out that processes in thermo-
dynamics serve a dual role. The first is, obviously,
the alteration of the state of a physical system. The
second, and more important role, is their use to estab-
lish relationships between and among states, and it is
precisely this function that we have been intensively
investigating in our discussion of the relations
$\overset{\beta}{\sim}$, $=_\beta$, $\leq_\beta$, and $<_\beta$. The capability of interrelating
points by processes is what makes processes so impor-
tant in thermodynamics. The importance of processes
is reflected in the next axiom, which deals exclusively
with processes and their properties. This axiom will
make assumptions about subsets of simple processes
which will just be adequate for the generation of alge-
braic structure on the set of thermodynamic state Σ.
In particular, the axiom will introduce adiabatic pro-
cesses which will ultimately become the adiabatic pro-

cesses of thermodynamics. However, initially, these processes will not be connected with heat and will be regarded merely as a distinguished subset with some assumed properties.

Axiom I.33: The set of all simple processes Π contains the subsets $\mathcal{O} \subset \mathcal{A} \subset \mathcal{P} \subset \Pi$, where $\mathcal{P}$ is the collection of all simple physical processes, $\mathcal{A}$ is the collection of all simple, physical, adiabatic processes and $\mathcal{O}$ is the collection of all simple, physical null processes. The subsets satisfy the following conditions: (1) $p \in \mathcal{P}$ or $p^* \in \mathcal{P}$ $\blacktriangledown$ $p \in \Pi$, (2) $(\mathcal{P} - \mathcal{A}) \cap \mathcal{A}^* = \emptyset$ and $\mathcal{A} \neq \mathcal{A}^*$, (3) for each $x \in \Sigma$ $\exists$ some $p \in \mathcal{O}$ such that $F(p) = (x, x)$, (4) for each x, $x' \in \Sigma$ $\exists$ some $\mathcal{P}$-process P such that $x \xrightarrow{P} x'$ or $x' \xrightarrow{P} x$, (5) if $\exists$ $\mathcal{A}$-processes P and $\overline{P}$ such that $x \xrightarrow{P} x'$ and $x' \xrightarrow{\overline{P}} x$ then, $\exists$ an $(\mathcal{A} \cap \mathcal{A}^*)$-process $\widetilde{P}$ such that $x \xrightarrow{\widetilde{P}} x'$, and (6) if $\exists$ $\mathcal{A}$-processes P' and P'' such that $(x' \xrightarrow{P'} x$ and $x'' \xrightarrow{P''} x)$ or $(x \xrightarrow{P'} x'$ and $x \xrightarrow{P''} x'')$ then, $\exists$ an $\mathcal{A}$-process P such that $x' \xrightarrow{P} x''$ or $x'' \xrightarrow{P} x'$.

In this paper the adjective physical has been applied to processes for the first time in Axiom I.33, and it might be well to comment on the connotation of that word in this connection. The elements of $\mathcal{P}$ and $\mathcal{P}$-processes should be regarded as the mathematical analogs of processes which can actually be carried out, in some sense, in the real world. Concomitantly, this interpretation implies that perhaps not all processes are realizable in the real world. As a matter of fact part (1) of the axiom assumes that either a simple process or its reverse is realizable, and part (4) makes the eminently reasonable assertion that any pair of states can be linked by some realizable process. An observant reader may have noticed that a statement equivalent to part (3) of the axiom was used in connection with equivalence relations and partial order relations, while the equivalents of parts (5) and (6) appeared in connection with partial order properties.

Definition I.34: An $\mathcal{A}$-process is called an adiabatic process, while a $\mathcal{P}$-process is said to be a physical process. An $\mathcal{A} \cap \mathcal{A}^*$-process is called a reversible adiabatic process and a $\mathcal{P} \cap \mathcal{P}^*$-process is said to be a reversible process.

Immediate consequences of Axiom I.33 are contained in the next theorem.

Theorem I.35: $\emptyset \neq \mathcal{O} \subset \mathcal{A} \cap \mathcal{A}^* \subset \mathcal{P} \cap \mathcal{P}^* \subset \mathcal{P} \subset \mathcal{P} \cup \mathcal{P}^* = \Pi$. Let $p \in \mathcal{A}$ and P an $\mathcal{A}$-process. Then p and P are reversible iff they are reversible in $\mathcal{A}$.

Proof: Suppose $p \in \Pi$; then, by Axiom I.33(1), $p \in \mathcal{P}$ or $p^* \in \mathcal{P}$, and, by Theorem I.8(2), $p \in \mathcal{P}^*$; hence $p \in \mathcal{P} \cup \mathcal{P}^*$ and thus $\Pi \subset \mathcal{P} \cup \mathcal{P}^*$. The converse $\mathcal{P} \cup \mathcal{P}^* \subset \Pi$ follows similarly. By Axiom I.33(3), $\mathcal{O} \neq \emptyset$ and, by Theorem I.9, $\mathcal{O} = \mathcal{O}^*$. Since $\mathcal{O} \subset \mathcal{A} \subset \mathcal{P}$, we have $\mathcal{O} = \mathcal{O} \cap \mathcal{O}^* \subset \mathcal{A} \cap \mathcal{A}^* \subset \mathcal{P} \cap \mathcal{P}^* \subset \mathcal{P} \cup \mathcal{P}^*$. Since $\mathcal{A} \cap \mathcal{A}^* \subset \mathcal{P} \cap \mathcal{P}^*$, it follows from Theorem I.8(4) and Theorem I.15(3) that if p and P are reversible in $\mathcal{A}$, they are reversible in $\mathcal{P} \cap \mathcal{P}^*$ and hence reversible. Conversely, suppose p and P are reversible. Then $p \in \mathcal{A} \Rightarrow p^* \in \mathcal{A}^*$ and p, $p^* \in \mathcal{P}$. But then $p^* \in \mathcal{P} \cap \mathcal{A}^* = \mathcal{A} \cap \mathcal{A}^*$ by Axiom I.33(2) and, by Corollary I.12, $p \in \mathcal{A} \cap \mathcal{A}^*$. Thus, p is reversible in $\mathcal{A}$. Since $\mathrm{Im}P \subset \mathcal{A}$, then $\mathrm{Im}P^* \subset \mathcal{A}^*$ and, since P is reversible, $\mathrm{Im}P^* \subset \mathcal{P}$ and, hence, $\mathrm{Im}P^* \subset \mathcal{P} \cap \mathcal{A}^* = \mathcal{A} \cap \mathcal{A}^* \Rightarrow \mathrm{Im}P \subset \mathcal{A} \cap \mathcal{A}^*$ and thus P is reversible in $\mathcal{A}$.

As a consequence of Definition I.34 and $\mathcal{P} \cap \mathcal{P}^* \subset \mathcal{P}$ we see that all reversible processes are assumed to be actually realizable in some sense. A sketch of the relationship among various kinds of simple processes is shown by the Venn diagram of Fig. 3.

Given the content of Axiom I.33, we can immediately utilize some of the previously derived results in order to introduce various equivalence relations on Σ. Thus, we have

Theorem I.36: The relations $\overset{A}{\sim}$, $=_A$, $\overset{P}{\sim}$, and $=_P$ are equivalence relations on Σ and, furthermore, (1) each equivalence class of $=_A (=_P)$ is a subset of some equivalence class of $\overset{A}{\sim} (\overset{P}{\sim})$ and (2) the relation $\overset{P}{\sim}$ possesses only one equivalence class, namely Σ itself.

Proof: Except for part (2) this theorem is a direct consequence of Axiom I.33 and Theorem I.18. To establish part (2) we know by Theorem I.34 that $\Pi = \mathcal{P} \cup \mathcal{P}^*$. Thus, it follows that every process is a $(\mathcal{P} \cup \mathcal{P}^*)$-process and by Axiom I.33(4) each pair of points in Σ is linked by some process and, hence, for each $x \in \Sigma$, $x \overset{P}{\sim} x'$ $\blacktriangledown$ $x' \in \Sigma$.

Definition I.37: If $x \overset{A}{\sim} x'$, then x and x' are said to be adiabatically equivalent. The equivalence classes of $\overset{A}{\sim}$ are called adiabatic components and denoted by Γ_λ, $\lambda \in \Lambda$ and Λ an index set. If $x =_A x'$, then x and x' are said to be adiabatically equal. The equivalence classes of $=_A$ are denoted by $[x]$, $x \in \Sigma$.

The adiabatic components represent rigorous generalizations of the sets $\beta, \beta', \beta'', \cdots$, introduced heuristically by Landsberg (1961, p. 31). As a matter of fact, in the next theorem we shall encounter subsets of the adiabatic components which themselves can be regarded as generalizations of Landsberg's $\beta, \beta', \beta'', \circ \circ \circ$.

Theorem I.38: The triplet $(\Sigma, \leqslant_A, =_A)$ is a nonbranching partially ordered set whose maximal chains partition the adiabatic components. For each $x \in \Sigma$, $[x]$ is subset of some maximal chain, and $x =_A x'$ iff $x \leqslant_A x'$ and $x' \leqslant_A x$ $\blacktriangledown$ x, $x' \in \Sigma$.

Proof: Combining Axiom I.33 with Theorem I.32 and Definition I.19 shows that $(\Sigma, \leqslant_A, =_A)$ is a partially ordered set and satisfies $x =_A x'$ iff $x \leqslant_A x'$ and $x' \leqslant_A x$. But by Axiom I.33(6) and Definition I.27 the relation $\leqslant_A$ is nonbranching and by Theorem I.30 the maximal chains partition Σ. Because $\mathcal{A} \cap \mathcal{A}^* \subset \mathcal{A} \subset \mathcal{A} \cup \mathcal{A}^*$, it fol-
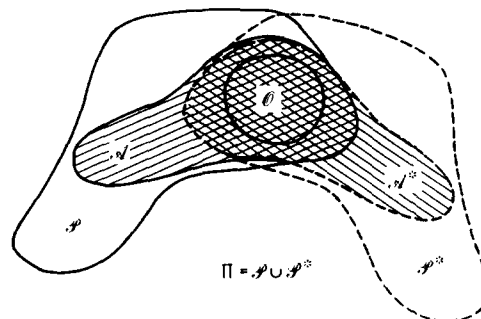
FIG. 3. Interrelationship among various kinds of simple processes.

$$\Sigma = \Gamma_1 \cup \Gamma_2 \cup \Gamma_3 \cup \Gamma_4$$
$$\Gamma_1 = C_1 \cup C_2 \cup C_3$$
$$C_1 = [x] \cup [x'] \cup [x'']$$
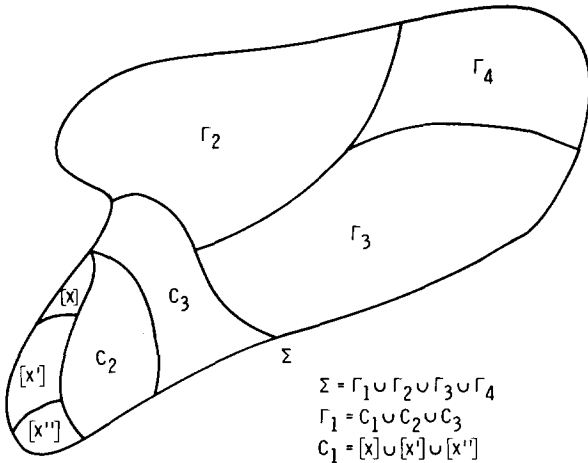
FIG. 4. An example of the partitioning of the set of thermo-dynamic states Σ by adiabatic processes. The Γ_i are adiabatic components, the C_i are maximal chains, and $[x]$, $[x']$, $[x'']$ are the equivalence classes of adiabatic equality.

lows that $x =_A x' \Rightarrow x \leqslant_A x' \Rightarrow x \overset{\triangle}{=} x'$. Thus, each equivalence class $[x]$ meets only one chain, each chain meets only one adiabatic component and the theorem is proved.

The maximal chains C of the nonbranching partial order $\leqslant_A$ can also be regarded as generalizations of Landsberg's $\beta, \beta', \beta'', \circ\circ\circ$. They can also be thought of as the equivalent of Boyling's[9] simple systems. See postulate IIa in Boyling (1972). An example of a situation that is compatible with Theorem I.38 is illustrated in the Venn diagram of Fig. 4. Here there are four adiabatic components but only the internal structure of Γ_1 is depicted. It contains three maximal chains and, again, only the internal structure of the maximal chain C_1 is pictured. It contains three equivalence classes of the relation $=_A$. The content of Theorem I.38 can be rephrased in terms of the equivalence classes $[x]$, $x \in \Sigma$.

Theorem I.39: Let $S = \{[x] \mid x \in \Sigma\}$ be the collection of equivalence classes of the relation $=_A$. If $=$ is the equivalence relation of ordinary equality in S and if $[x] \leqslant [x']$ iff $x \leqslant_A x'$, then $(S, \leqslant, =)$ is a nonbranching partially ordered set whose maximal chains partition S. Let $\Gamma = \{\Gamma_\lambda \mid \lambda \in \Lambda\}$ be the collection of adiabatic components of Σ and $S_\lambda = \{[x] \mid x \in \Gamma_\lambda\}$. Then (1) $\{S_\lambda \mid \lambda \in \Lambda\}$ is a partition of S, (2) the maximal chains of S partition S_λ, and (3) $C = \{[x] \mid x \in C, \ C$ a chain in $\Sigma\}$ is a maximal chain in S iff C is a maximal chain in Σ.

Proof: From the fact that $=_A$ is an equivalence relation it follows that $[x] = [x']$ iff $x =_A x'$. This, coupled with the definition of the relation $\leqslant$ and the fact that $\leqslant_A$ is a nonbranching partial order relation, establishes that $(S, \leqslant, =)$ is a nonbranching partially ordered set and hence its maximal chains partition S by Theorem I.30. Next we show that $\{S_\lambda \mid \lambda \in \Lambda\}$ is a partition of S. Obviously, $S_\lambda \neq \emptyset$ since $\Gamma_\lambda \neq \emptyset$ and $\cup_\lambda S_\lambda \subset S$. Suppose $[x] \in S$; then $x \in \Sigma$ and hence $x \in \Gamma_\lambda$ for some λ since Γ is a partition of Σ and hence $[x] \in S_\lambda$ for some λ. Thus, $S \subset \cup_\lambda S_\lambda$ and hence $S = \cup_\lambda S_\lambda$. Finally, suppose $S_\lambda \cap S_{\lambda'}$

$\neq \emptyset$. Then $\exists [x]$ such that $[x] \in S_\lambda$, $[x] \in S_{\lambda'}$ and then $x \in \Gamma_\lambda$ and $x \in \Gamma_{\lambda'}$; hence, $\Gamma_\lambda \cap \Gamma_{\lambda'} \neq \emptyset$ which is a contradiction. Thus, $\{S_\lambda \mid \lambda \in \Lambda\}$ is a partition. Now consider the set $C = \{[x] \mid x \in C, \ C$ a chain in $\Sigma\}$. Since $x \leqslant_A x'$ or $x' \leqslant_A x$ iff $[x] \leqslant [x']$ or $[x'] \leqslant [x]$ it follows that C is a chain in S iff C is a chain in Σ. Further, $C_1 = \{[x] \mid x \in C_1\} \subset C_2 = \{[x] \mid x \in C_2\}$ iff $C_1 \subset C_2$; hence, $C_1 = C_2$ iff $C_1 = C_2$ and thus C is a maximal chain in S iff C is a maximal chain in Σ. Since the maximal chains in S partition S, we need only show that if C is a maximal chain, then $C \subset S_\lambda$ for some λ. But $[x] \in S_\lambda$ iff $x \in \Gamma_\lambda$, and hence $C \subset S_\lambda$ iff $C \subset \Gamma_\lambda$. But $C \subset \Gamma_\lambda$ from Theorem I.38.

Corollary I.40: There exists a 1–1 correspondence between the maximal chains in S and the maximal chains in Σ.

Proof: Let $\{C\}$ be the collection of maximal chains in Σ and $\{C\}$ the collection of maximal chains in S. Define the relation $g: \{C\} \rightarrow \{C\}$ by $g(C) = C = \{[x] \mid x \in C\}$. By the proof of the previous theorem $C_1 = C_2$ iff $C_1 = C_2$ iff $g(C_1) = g(C_2)$. Thus, g is a 1–1 function. Further, if C is a maximal chain, then $C = \{x \mid [x] \in C\}$ is its preimage and hence g is onto.

The "mutual accessibility" classes of Boyling[9] and Buchdahl[14,15] and the "frontier sets" of Rastall[16] are analogous to my equivalence classes $[x]$ with the following clarification. My equivalence classes and the "mutual accessibility" classes are purely algebraic in origin and invoke only adiabatic accessibility. Rastall's[16] "froniter sets," on the other hand, involve not only adiabatic accessibility but topological assumptions as well (p. 2958, Axioms VI and VII together with the assumption that "frontier sets" are closed).

High on the list of important thermodynamic concepts is the notion of equilibrium states, yet, sometimes, this idea is poorly defined in theoretical treatments of thermodynamics. An unambiguous definition is

Definition I.41: Let C be a chain, not necessarily maximal in $(\Sigma, \leqslant_A, =_A)$. An equilibrium state of C is a maximal element of C.

Three things should be observed about this definition. First, an equilibrium state is defined with respect to a chain and unless one is prepared to specify the chain one cannot speak about an equilibrium state. Second, a particular chain may not have an equilibrium state since we are not assured of the existence of maximal elements in a chain. Third, an equilibrium state, if it exists, may not be unique. This is shown in the next theorem.

Theorem I.42: Let C be a chain in $(\Sigma, \leqslant_A, =_A)$ and e an equilibrium state of C. If $e' \in [e]$, then e' is an equilibrium state of C.

Proof: Suppose e is an equilibrium state of C. Then, since e is maximal, $e \leqslant_A x \Rightarrow x =_A e$ for $x \in C$. Further, $e' \in [e] \Longleftrightarrow e =_A e'$ iff $e \leqslant_A e'$ and $e' \leqslant_A e$. Suppose $e' \leqslant_A x$ and $x \neq_A e'$. By the transitivity of $\leqslant_A$ and $=_A$ we know that $e' \leqslant_A x$ and $e \leqslant_A e' \Rightarrow e \leqslant_A x \Rightarrow x =_A e$ and, since $e =_A e'$, $\Rightarrow x =_A e'$ which is a contradiction. Thus, $e' \leqslant_A x \Rightarrow x =_A e'$ and e' is an equilibrium state of C.

There are two ways in which one can arrive at a unique equilibrium state if equilibrium states exist.

One would certainly have a unique equilibrium state if the cardinality of $[e]$ were one. Alternately, and more generally, one could impose additional conditions on the states in C which would permit one to eliminate all members of $[e]$ except one. This is equivalent to assuming the axiom of choice, that is, endowing ourselves with the ability to select one element from each of the equivalence classes in C. This, in fact, represents a generalization of what is conventionally done in thermodynamics. There one encounters statements such as (1) equilibrium is the state of maximum entropy for fixed internal energy and fixed volume or (2) equilibrium is the state of minimum Gibbs free energy for fixed pressure and fixed temperature. The axiom of choice is a standard mathematical tool and is equivalent to Zorn's lemma which appeared as Lemma I.25.

Theorem I.43: Axiom of Choice: Let $\{A_\alpha \mid \alpha \in \Delta\}$ be a nonempty collection of pairwise disjoint sets. Then **∃** a set A such that $A_\alpha \cap A$ has precisely one element in A_α for each $\alpha \in \Delta$. Equivalently, let $\{A_\alpha \mid \alpha \in \Delta\}$ be a nonempty collection of sets, disjoint or not. Then **∃** a function $\gamma : \Delta \rightarrow \cup_{\alpha \in \Delta} A_\alpha$, called a choice function, such that $\gamma(\alpha) \in A_\alpha$ for each $\alpha \in \Delta$.

The axiom of choice is intimately connected with the extension of the definition of the Cartesian product from the case of a finite number of factors, e.g., $\Sigma_1 \times \Sigma_1$, to the case of an infinite number of factors. This standard extension is given as part of the next definition.

Definition I.44: Let $\{A_\alpha \mid \alpha \in \Delta\}$ be a nonempty collection of sets, disjoint or not. Then the set of all choice functions is called the Cartesian product of $\{A_\alpha \mid \alpha \in \Delta\}$ and designated by $\Pi_{\alpha \in \Delta} A_\alpha$. For $\beta \in \Delta$ the βth projection map, π^β, is the function $\pi^\beta : \Pi_{\alpha \in \Delta} A_\alpha \rightarrow A_\beta$ defined by $\pi^\beta(\gamma) = \gamma(\beta) \equiv \gamma^\beta$ **∀** $\gamma \in \Pi_{\alpha \in \Delta} A_\alpha$.

Theorem I.45: Let C be a chain in $(\Sigma, \leqslant_A, =_A)$ with a maximal element and $\Pi[x]$ the Cartesian product of the distinct equivalence classes in C. Then for each $\gamma \in \Pi[x]$ **∃** a unique equilibrium state of C.

Proof: Since γ is a choice function, the discussion preceding the axiom of choice applies.

The fundamental structure of thermodynamics is now complete and arises from only two axioms, namely Axiom I.1 and Axiom I.33. This has been accomplished solely with the ideas of states and adiabatic processes without any mention of heat, work, the first law, temperature, entropy, etc. This is not to say that these ancillary trappings are unimportant for the applications of thermodynamics but only that the fundamentals of thermodynamics in no way depend upon them. The use of the word adiabatic in the preceding development might give one the erroneous impression that heat has been implicitly used in the concept of an adiabatic process. The fact is that adiabatic processes have only been required to possess the properties listed in Axiom I.33 and any subset of $\mathcal{P}$ with the requisite properties will suffice to generate a thermodynamics. The choice of a name for the subset, although operationally significant, is mathematically insignificant. The general thermodynamics that I have developed specializes to a form that is obviously similar to conventional thermodynamics if one assumes that Σ itself is the only

adiabatic component and is also the only maximal chain. In effect, conventional thermodynamics is the restrictions of the general structure to a maximal chain.

Of course, any theory which aspires to be called thermodynamics must eventually come to grips with such things as entropy and the first law. The definition of entropy presents no problem.

Definition I.46: Let C be a chain in $(\Sigma, \leqslant_A, =_A)$. An empirical entropy for C is an order homomorphism from C onto a subset of the reals with the usual order. That is, φ is an empirical entropy iff $\varphi : C \rightarrow R$ is a function such that (1) $\varphi(x_1) \leqslant \varphi(x_2)$ iff $x_1 \leqslant_A x_2$ and (2) $\varphi(x_1) = \varphi(x_2)$ iff $x_1 =_A x_2$ **∀** $x_1, x_2 \in C$.

Strictly speaking condition (2) of this definition is redundant because it follows from condition (1) and the property $x =_A x'$ iff $x \leqslant_A x'$ and $x' \leqslant_A x$ of Theorem I.38. Thus, $x_1 =_A x_2$ iff $x_1 \leqslant_A x_2$ and $x_2 \leqslant_A x_1$ iff $\varphi(x_1) \leqslant \varphi(x_2)$ and $\varphi(x_2) \leqslant \varphi(x_1)$ iff $\varphi(x_1) = \varphi(x_2)$. The need for restricting the definition of empirical entropy to a chain rather than Σ itself is obvious when one recalls that the reals with their usual order are a chain. But a chain can only reflect the order properties of another chain if at all. The desire to retain the notion of entropy is what dictated the requirement that $\leqslant$ be a nonbranching partial order in the first place. As we have seen, a nonbranching partial order partitions a set into maximal chains and hence each element of a set is in some maximal chain. The existence of maximal chains does not guarantee the existence of entropies but at least it makes it reasonable to attempt their construction. A substantial portion of the remainder of the paper will be devoted to the construction of entropy functions. Definition I.46 presents an unambiguous answer to the oft repeated question "what is entropy?" Simply put, it is a real-valued function which *reflects* the order, in a chain, induced by the adiabatic processes on Σ. It decidely is *not* a real-valued function which *establishes* an order because every real-valued function induces some order on its domain by virtue of the usual order on the reals and this order, in general, will not be the same as the adiabatic order. The entropy is now recognized as a purely mathematical, rather than a physical concept and this may account for some of the conceptual difficulties surrounding the entropy in thermodynamics. Clearly, the concept of entropy has no extension to all states X unless it possesses an extension to Σ and, of course, it can only be extended to all thermodynamic states if Σ itself is a maximal chain. Thus it is of prime physical importance to establish experimentally the extent of the maximal chains in Σ. The definition of entropy makes it clear that an entropy cannot be unique, for if φ is an entropy, $k_1 > 0$ and k_2 real numbers, then the function φ_{k_1, k_2} defined by $\varphi_{k_1, k_2}(x) = k_1 \varphi(x) + k_2$ is also an entropy. If we insist that an entropy exist for a chain C, something we have not yet done, then we must accept the fact that C cannot contain more equivalence classes than the cardinality of the continuum. This is shown in the next theorem.

Theorem I.47: Let φ be an empirical entropy for a chain C in $(\Sigma, \leqslant_A, =_A)$ and $\mathcal{C} = \{[x] \mid x \in C\}$. Then $\mathcal{C}$ is order isomorphic to a subset of the reals.

Proof: Define $\phi : \mathcal{C} \rightarrow R$ by $\phi([x]) = \varphi(x)$. Then $\phi([x_1])$

$-\phi([x_2]) = 0$ iff $\varphi(x_1) - \varphi(x_2) = 0$ iff $x_1 =_A x_2$ iff $[x_1]$ $= [x_2]$. Thus, ϕ is a 1–1 function. Further, this also demonstrates that the second part of Definition I.23 is satisfied. Similarly, $\phi([x_1]) - \phi([x_2]) \leqslant 0$ iff $\varphi(x_1)$ $- \varphi(x_2) \leqslant 0$ iff $x_1 \leqslant_A x_2$ iff $[x_1] \leqslant [x_2]$.

Finally, we have reached the point where it becomes appropriate to introduce the first law of thermodynamics.

Axiom I.48: The First Law: Let X be the set of all physical states and Π the collection of all simple processes on the collection of thermodynamic states Σ. Then $\exists$ real-valued functions $Q : \Pi \to R$, $W : \Pi \to R$, and $\Delta u : X \times X \to R$ such that (1) $Q(p) + W(p) = \Delta u(x, x')$ $\blacktriangledown$ $p \in \Pi$ and x, $x' \in \Sigma$ such that $F(p) = (x, x')$, (2) $Q(p) + Q(p^*) = 0$, $W(p) + W(p^*) = 0$, $\blacktriangledown$ $p \in \Pi$, and (3) $\Delta u(x, x') + \Delta u(x', x'') = \Delta u(x, x'')$ $\blacktriangledown$ x, x', $x'' \in X$. The values $Q(p)$ and $W(p)$ are called the heat and work, respectively, of a simple process p, while $\Delta u(x, x')$ is the internal energy increment between x, $x' \in X$.

A few explanatory comments about Axiom I.48 are in order. The domain of both Q and W is taken to be Π and condition (2) is interpretable as a property of heat and work. However, we could equally well have chosen the domain to be the set of simple, physical processes P. Then condition (2) would be interpreted as the extension of heat and work from P to Π. The reason for this is that, by Axiom I.33, (1) $p \in P$ or $p^* \in P$ $\blacktriangledown$ $p \in \Pi$. Also the domain of Δu was taken to be $X \times X$ rather than $\Sigma \times \Sigma$ in order to imply that Δu is significant even for nonthermodynamic states. Of course the thermodynamic applications of Δu will only involve its restriction to $\Sigma \times \Sigma$.

Theorem I.49: The restriction of Δu to $\Sigma \times \Sigma$ is a skew-symmetric function, that is, $\Delta u(x, x) = 0$ and $\Delta u(x, x') = -\Delta u(x', x)$ $\blacktriangledown$ $x, x' \in \Sigma$. The relation $u_{x'} : \Sigma \to R$ defined by $u_{x'}(x) = \Delta u(x', x)$ $\blacktriangledown$ $x \in \Sigma$ is a function $\blacktriangledown$ $x' \in \Sigma$ and is called the internal energy relative to x'.

Proof: Let $p \in \Pi$ and $F(p) = (x, x')$. From Axiom I.48 it follows that $0 = Q(p) + W(p) + Q(p^*) + W(p^*) = \Delta u(x, x')$ $+ \Delta u(x', x) = \Delta u(x, x)$. To show that $u_{x'}$ is a function observe that $u_{x'}(\bar{x}) - u_{x'}(x) = \Delta u(x', \bar{x}) - \Delta u(x', x) = \Delta u(x', \bar{x})$ $+ \Delta u(x, x') = \Delta u(x, \bar{x})$. Thus, if $x = \bar{x}$, then $u_{x'}(x) = u_{x'}(\bar{x})$ since $\Delta u(\bar{x}, \bar{x}) = 0$.

We shall find it convenient to postulate that certain simple processes have zero heat or zero work.

Axiom I.50: (1) If $p \in A$, then $Q(p) = 0$, and (2) if $p \in O$, then $W(p) = 0$.

Note that since $O \subset A$ we have the fact that simple null processes are both zero heat and zero work processes. Further, the converses of the statements in Axiom I.50 are not generally true. For example, $Q(p) = 0$ does not imply $p \in A$. In fact we have

Theorem I.51: $Q(p) = 0$ $\blacktriangledown$ $p \in A \cup A^*$.

Proof: If $p \in A \cup A^*$, then $p \in A$ or $p \in A^*$ and, by Theorem I.8, $p \in A$ or $p^* \in A$ and hence $Q(p) = 0$ or $Q(p^*) = 0$. But, by Axiom I.48 (2a) $Q(p) = -Q(p^*)$.

It now is a very simple matter to extend the ideas of

heat and work from simple processes to processes. This extension is carried out in the next theorem.

Theorem I.52: Define relations from the collection of all processes to the reals by $Q(P) \equiv \sum_n Q(p_n)$ and $W(P) \equiv \sum_n W(p_n)$, where P is a process such that $P : N \to \Pi$ with $P(n) = p_n$ and $x \xrightarrow{P} x'$. Then the extensions of Q and W are real-valued functions on the collection of all processes. Further, (1) $Q(P) + W(P) = \Delta u(x, x')$, (2) $Q(P) + Q(P^*) = 0$, $W(P) + W(P^*) = 0$, (3) if P is an $(A \cup A^*)$-process, then $Q(P) = 0$, and (4) if P is a null process $Q(P) = 0 = W(P)$.

Proof: Simple null processes have zero heat and work by Axiom I.50. Since a process contains at most a finite number of nonnull simple processes, the sums $Q(P)$ and $W(P)$ are finite and hence convergent. Suppose $P = \bar{P}$. Then $p_n = \bar{p}_n$ $\blacktriangledown$ $n \in N$ and therefore $Q(p_n) = Q(\bar{p}_n)$ and hence $Q(P) = \sum_n Q(p_n) = \sum Q(\bar{p}_n) = Q(\bar{P})$. Similarly, $W(P) = W(\bar{P})$ and thus the extensions are real-valued functions. The second part is easily established since by Definition I.14, if P is a process of length n_0, then for its reverse P^* we have $P^*(n) = p^*_{n_0+1-n}$ for $n \leqslant n_0$ and $P^*(n) \in O$ for $n > n_0$. Thus,

$$Q(P^*) = \sum_n Q[P^*(n)] = \sum_{n=1}^{n_0} Q(p^*_{n_0+1-n})$$

$$= -\sum_{n=1}^{n_0} Q(p_{n_0+1-n}) = -\sum_{n=1}^{n_0} Q(p_n) = -\sum_n Q(p_n) = -Q(P).$$

In similar fashion $W(P) + W(P^*) = 0$. The third part follows from Theorem I.51 and the definition of an $(A \cup A^*)$-process. To prove (1), suppose P is a process of length n_0 linking x and x'. Then

$$Q(P) + W(P) = \sum_n [Q(p_n) + W(p_n)]$$

$$= \sum_{n=1}^{n_0} [Q(p_n) + W(p_n)] = \sum_{n=1}^{n_0} \Delta u(a_n, b_n),$$

where $F(p_n) = (a_n, b_n)$. But, from Definition I.13, $b_n = a_{n+1}$ and using Axiom I.48 (3) $Q(P) + W(P)$ $= \sum_{n=1}^{n_0} \Delta u(a_n, a_{n+1}) = \Delta u(a_1, a_{n_0+1}) = \Delta u(a_1, b_{n_0}) = \Delta u(x, x')$. Finally, to establish (4), suppose P is a null process. Then $p_n \in O$ $\blacktriangledown$ $n \in N$ and, from Axiom I.50, $Q(P) = 0$ $= W(P)$.

Heat, like work and the first law, has played no role in developing the fundamental structure of thermodynamics, as has already been pointed out. But its role in thermodynamics can be quite easily established in spite of the fact that no explicit definition for heat has been given. It turns out that heat offers us a means for deciding whether or not two points x, x' in a maximal chain are adiabatically equal, $x =_A x'$. This then leads to a possible construction method for empirical entropies. We start with a simple theorem which will lay the groundwork for establishing these assertions.

Theorem I.53: x, $x' \in \Sigma$ and P a P-process such that $x \xrightarrow{P} x'$. If $x =_A x'$, then $\exists$ a P-process $\bar{P}$ such that $x \xrightarrow{E} x$, $Q(\bar{P}) + W(\bar{P}) = 0$, and $Q(\bar{P}) = Q(P)$.

Proof: By Theorem I.38, $x =_A x'$ iff $x \leqslant_A x'$ and $x' \leqslant_A x$, and hence $\exists$ A-processes P_1, P_2 such that $x \xrightarrow{P_1} x'$ and $x' \xrightarrow{P_2} x$. Consider the process $\bar{P}$ such that $x \xrightarrow{P} x' \xrightarrow{P_2} x$, that is, $\bar{P}(n) = P(n)$ for $n \leqslant n_0$ and $\bar{P}(n)$

$= P_2(n - n_0)$ for $n > n_0$, where n_0 is the length of P. Then $x \xrightarrow{\bar{P}} x$ and thus $Q(\bar{P}) + W(\bar{P}) = \Delta u(x, x) = 0$ and $\bar{P}$ is a P-process. Also $Q(\bar{P}) = Q(P) + Q(P_2) = Q(P)$ since $Q(P_2) = 0$.

This theorem asserts that for any pair of adiabatically equal points connected by some physical process P there exists another physical process $\bar{P}$ whose sole effect is the complete interconversion of heat and work. This presents no difficulties if $Q(P) \leq 0$ for then $W(\bar{P}) = - Q(P) \geq 0$ and this corresponds to the conversion of work to heat. If, however, $Q(P) > 0$, then $W(\bar{P}) < 0$ and heat has been completely converted to work without a compensating change in the system. This violates Kelvin's statement[2,3] of the second law and is a result at variance with experience. This suggests that we adopt

Axiom I.54: $x =_A x'$ iff $Q(P) \leq 0$ ∀ P-processes P such that $x \xrightarrow{P} x'$ or $x' \xrightarrow{P} x$.

This axiom plays the role of a second law because it is designed to prevent the occurrence of the same situations which are prohibited by the second law. One immediate consequence of Axiom I.54 is $Q(P) \leq 0$ ∀ P-processes such that $x \xrightarrow{P} x$ because $x =_A x$. The combination of Axiom I.54 and Theorem I.52(2) shows that if P is a reversible process linking x and x' where $x =_A x'$, then $Q(P) = 0$. Another immediate and important consequence of Axiom I.54 is

Theorem I.55: $x <_A x'$ iff $x \leq_A x'$ and $Q(P) > 0$ for some P-process P such that $x \xrightarrow{P} x'$ or $x' \xrightarrow{P} x$.

Proof: From Theorem I.21, $x <_A x'$ iff $x \leq_A x'$ and $x \neq_A x'$. But the contraposition of Axiom I.54 is $x \neq_A x'$ iff $Q(P) > 0$ for some P-process P such that $x \xrightarrow{P} x'$ or $x' \xrightarrow{P} x$.

The significance of Axiom I.54 and Theorem I.55 cannot be overestimated for they take two relations, $=_A$ and $<_A$, whose definitions involve the concept of reversible processes, and recast them into a form which is wholly independent of the notion of reversible processes. The idea of reversibility has been replaced by the idea of the heat of a process. This enables us to regard reversible processes as mere mathematical constructs rather than true physical reality if we wish. Such a decision need not be made here. It is important to point out that, while the first law was used in the proof of Theorem I.53, both Axiom I.54 and Theorem I.55 are independent of the first law. The only purpose for Theorem I.53 was to provide motivation for Axiom I.54. Both Axiom I.54 and Theorem I.55 become dependent upon the first law only if the first law is used to define heat, that is, if heat is defined in terms of work and the internal energy increment.

Definition I.56: For any pair of points x, $x' \in \Sigma$ define a subset of the reals by $S(x, x') = \{Q(P) > 0 \mid x \xrightarrow{P} x'$ or $x' \xrightarrow{P} x$, $\mathrm{Im}P \subset P\} \subset R$. The greatest lower bound of $S(x, x')$, $\inf[S(x, x')]$, is called the minimum heat for x and x' if it exists.

The next few theorems will utilize Definition I.56 in an attempt to construct empirical entropies from heat measurements alone.

Theorem I.57: Let C be some chain in $(\Sigma, \leq_A, =_A)$ and $\varphi : C \times C \to R$ be a relation defined by $\varphi(x, x') = 0$ if

$S(x, x')$ is a null set, $\varphi(x, x') = \inf[S(x, x')]$ if $x <_A x'$ and $\varphi(x, x') = - \inf[S(x, x')]$ if $x' <_A x$. The relation φ is a skew-symmetric function and satisfies (1) $x =_A x'$ $\Rightarrow \varphi(x, x') = 0$ and (2) $x \leq_A x' \Rightarrow \varphi(x, x') \geq 0$. If $\inf[S(x, x')] \in S(x, x')$ ∀ x, $x' \in C$ and $x \neq_A x'$, then (3) $\varphi(x, x') = 0$ iff $x =_A x'$ and (4) $\varphi(x, x') \geq 0$ iff $x \leq_A x'$.

Proof: By Axiom I.54, $x =_A x'$ iff $S(x, x')$ is null $\Rightarrow \varphi(x, x') = 0$ which is (1). If $S(x, x')$ is not null, then $\inf[S(x, x')] \geq 0$. Hence, $x \leq_A x'$ iff $x <_A x'$ or $x =_A x'$ $\Rightarrow \varphi(x, x') \geq 0$ or $\varphi(x, x') = 0 \Rightarrow \varphi(x, x') \geq 0$ which is (2). To establish the skew-symmetry of φ we have, if $x =_A x'$, then $\varphi(x, x') + \varphi(x', x) = 0 + 0 = 0$ while, if $x <_A x'$, then $\varphi(x, x') + \varphi(x', x) = \inf[S(x, x')] - \inf[S(x, x')] = 0$ and similarly for $x' <_A x$. To show φ is a function we must show that $(x, x') = (\bar{x}, \bar{x}') \in C \times C \Rightarrow \varphi(x, x') = \varphi(\bar{x}, \bar{x}')$. But $(x, x') = (\bar{x}, \bar{x}')$ iff $x = \bar{x}$ and $x' = \bar{x}' \Rightarrow x =_A \bar{x}$ and $x' =_A \bar{x}'$. If $x =_A x'$, then $\bar{x} =_A x =_A x' =_A \bar{x}'$ and conversely. Thus, $x =_A x'$ iff $\bar{x} =_A \bar{x}'$ and then $\varphi(x, x') = 0 = \varphi(\bar{x}, \bar{x}')$. If $x \neq_A x'$ and $(x = \bar{x}$ and $x' = \bar{x}')$, then $x \xrightarrow{P} x' \Rightarrow \bar{x} \xrightarrow{P} \bar{x}'$. But then $S(x, x') = S(\bar{x}, \bar{x}')$ and again $\varphi(x, x') = \varphi(\bar{x}, \bar{x}')$. This establishes that φ is a function. To establish the last two parts of the theorem suppose, $\inf[S(x, x')] \in S(x, x')$. Then by the definition of S it follows that $\inf[S(x, x')] > 0$ and hence $\varphi(x, x') \neq 0$ iff $x \neq_A x'$. Thus, $\varphi(x, x') = 0 \Rightarrow x =_A x'$, which combined with (1) establishes (3). Further, $\varphi(x, x') > 0 \Rightarrow x <_A x'$, and then $\varphi(x, x') \geq 0 \Rightarrow x \leq_A x'$; this combined with (2) establishes establishes (4).

Even though we have not yet produced an empirical entropy, we have now reached the point where it becomes possible to write down conditions which are sufficient to guarantee the existence of an empirical entropy. The satisfaction of these conditions will also permit us to exhibit a number of empirical entropies.

Theorem I.58: Let C be some chain in $(\Sigma, \leq_A, =_A)$, $S(x, \bar{x})$ the subsets of the reals given in Definition I.56, φ the function of Theorem I.57, $x' \in C$, and $\varphi_{x'} : C \to R$ a map defined by $\varphi_{x'}(x) = \varphi(x', x)$. If ∀ x, $\bar{x} \in C$, we have (1) $\varphi(x, \bar{x}) \leq \varphi(x, x') + \varphi(x', \bar{x})$ and (2) $\inf[S(x, \bar{x})] \in S(x, \bar{x})$, then (3) $\varphi_{x'}$ is an empirical entropy called the empirical entropy relative to x' and (4) $\varphi(x, x') + \varphi(x', \bar{x}) = \varphi(x, \bar{x})$ ∀ x, $\bar{x} \in C$.

Proof: From (1) we have $\varphi(\bar{x}, x) \leq \varphi(\bar{x}, x') + \varphi(x', x)$. Multiplying by $(- 1)$ and using the skew-symmetry of φ leads to $\varphi(x, \bar{x}) \geq \varphi(x', \bar{x}) + \varphi(x, x') \geq \varphi(x, \bar{x})$, where the last inequality is just (1). Hence, $\varphi(x, \bar{x}) = \varphi(x, x') + \varphi(x', \bar{x})$, and the last part of the theorem has been proven. From this it follows that $\varphi_{x'}(\bar{x}) - \varphi_{x'}(x) = \varphi(x', \bar{x}) - \varphi(x', x) = \varphi(x, \bar{x})$. Hence, by the skew-symmetry of φ, it follows that $x = \bar{x} \Rightarrow \varphi_{x'}(x) = \varphi_{x'}(\bar{x})$ and hence $\varphi_{x'}$ is a function. Further, $\varphi_{x'}(\bar{x}) \leq \varphi_{x'}(x)$ iff $\varphi(x, \bar{x}) \leq 0$ iff $\varphi(\bar{x}, x) \geq 0$ iff $\bar{x} \leq_A x$ and $\varphi_{x'}(\bar{x}) = \varphi_{x'}(x)$ iff $\varphi(x, \bar{x}) = 0$ iff $\bar{x} =_A x$ and then, by Definition I.46, $\varphi_{x'}$ is an empirical entropy.

Suppose that there exists some $x' \in C$ which satisfies the conditions of Theorem I.58. Then by the skew-symmetry of $\varphi(x, x')$ it follows that $\varphi_{x'}(x') = \varphi(x', x') = 0$ and hence x' is a state of zero entropy, as is any other state which is adiabatically equal to it. Clearly, no particular significance can be attached to this result unless x' is unique in some sense, that is, x' itself

possesses some intrinsic significance. But the only possible way that an element of a chain could be intrinsically significant would be for it to be the largest or smallest element of the chain. These elements, if they exist, are unique but they need not exist. Intuitively one feels that a maximal chain, in thermodynamics, cannot possess a largest element. This certainly is conjecture and not proof. However, granting this intuitive feel, it then seems that the third law of thermodynamics can be regarded as an assertion that a maximal chain possesses a smallest element and that this element satisfies the conditions contained in Theorem I.58. We shall not pursue this further, but instead look at the situation where more than one point satisfies the conditions of Theorem I.58 and try to determine the relationships among the various entropies.

Corollary I.59: Suppose x', $x'' \in C$ both satisfy the conditions of Theorem I.58. Then $\blacktriangledown x \in C$ it is true that $\varphi_{x'}(x) - \varphi_{x''}(x) = \varphi(x', x'')$ and (1) $\varphi(x', x'') = 0$ iff $x' = {}_{A}x''$ and (2) $\varphi(x', x'') \geq 0$ iff $x' \leq {}_{A}x''$.

Proof: From the definitions of $\varphi_{x'}$ and $\varphi_{x''}$ it follows that $\varphi_{x'}(x) - \varphi_{x''}(x) = \varphi(x', x) - \varphi(x'', x) = \varphi(x', x) + \varphi(x, x'')$. Since x'' satisfies the conditions of Theorem I.58, it follows from Theorem I.58(4) that $\varphi(x, x'') + \varphi(x'', \bar{x}) = \varphi(x, \bar{x}) \blacktriangledown x, \bar{x} \in C$. Let $\bar{x} = x'$ and use the skew-symmetry of φ to get $\varphi(x, x'') + \varphi(x', x) = \varphi(x', x'')$. The remainder of the corollary can be proven by setting $x = x'$ in $\varphi(x, x'') = 0$ iff $x = {}_{A}x''$ and $\varphi(x, x'') \geq 0$ iff $x \leq {}_{A}x''$, which come from Theorem I.57(3, 4).

All that is now required is the postulate that the conditions of Theorem I.58 are actually satisfied. Such a postulate is unwise on two counts. First, the existence of an empirical entropy is not a requisite of thermodynamics and, second, the conditions of Theorem I.58 have not yet been subjected to an experimental test. There are some highly significant aspects of the foregoing discussion of entropy. The discussion is couched solely in terms of heat and physical processes. Nowhere is there a mention of the first law; if one can define heat without recourse to the first law, then entropy possesses an existence which is independent of the first law. Nowhere is there a need to come to grips with reversible processes and their experimental significance. One need know about adiabatic processes only to the extent that they determine the maximal chains which partition Σ. My reliance on adiabatic processes alone contrasts with the work of Buchdahl[14] (1958, 1962), who used not only adiabatic processes but also zero work processes, which he called isometric processes. Similarly, Rastall[16] employed adiabatic processes and a slight generalization of the isometric processes, which he called anergic processes. These authors used isometric and anergic processes to construct entropies. This could also be done within the framework of the theory constructed here. It would only be necessary to postulate the existence of such a set of processes and assume that these processes suitably link the equivalence classes $[x]$ of a maximal chain. This procedure has the disadvantage that one must characterize these processes sufficiently so that they can be recognized experimentally. Additionally, one must then verify that any postulated linking properties are realized

experimentally. To me it seems preferable to construct thermodynamics from only one class of processes, namely, the adiabatic processes.

All of the thermodynamics developed so far has dealt exclusively with the abstract, and hence undefined, concepts of states and processes connecting these states. An obvious advantage of this abstract approach is that it gives the experimentalist considerable latitude in selecting the real world counterparts of the abstract quantities. Any attempt to be less abstract, and thus more specific, will limit the choices available to the experimentalist. The move from the abstract to the concrete is desirable if it leads to a characterization of abstract notions in terms of ideas which are more easily interpretable in the real world. Since our characterization of phenomena in the real world is predicated on measurements, it is natural to attempt to interpret states and processes in terms of measurements. This is precisely what will be done and the opening move in that direction is a definition of measurement.

Definition I.60: Let Y be a subset of the set of all physical states X. A measurement m is a real-valued function on X, $m : X \rightarrow R$, and its restriction to Y, $m | Y$, is said to be a measurement on Y. A set of measurements $M = \{m^{\alpha} | Y \mid m^{\alpha} : X \rightarrow R, \alpha \in \Delta\}$ on Y is said to be a measurement set for Y iff it separates the points of Y, that is, $x \neq x' \in Y \Rightarrow m^{\alpha}(x) \neq m^{\alpha}(x')$ for some $\alpha \in \Delta$. A measurement set for Y is said to be a coordinate system for Y iff it is minimal, that is, M_0 is a coordinate system for Y iff M is a measurement set for Y and $M \subset M_0 \Rightarrow M = M_0$. The dimension of a coordinate system is its cardinality and its elements are called coordinates of Y.

A measurement is by definition a real-valued function; however, a real-valued function need not be a measurement since it is the experimentalist who decides which real-valued functions qualify as measurements. An alternative, but equivalent, definition of a coordinate system would be to say that a measurement set is a coordinate system for Y iff every proper subset does not separate the points of Y. In general, a given set Y could possess many coordinate systems and there is no *a priori* reason for all coordinate systems to have the same dimensionality. Obviously, the dimensionality will strongly depend on the character of the measurements. From an experimental point of view it is essential that the dimensionality be finite because there is no way to make an infinite number of measurements. Since the cardinal numbers are linearly ordered, it makes mathematical sense to speak of coordinate systems with minimal dimensionality. However, as a practical matter there is no way to know that a given system has this property since this requires comparing its dimensionality with the dimensionality of all other coordinate systems.

Theorem I.61: Let $Y \subset X$, $M = \{m^{\alpha} | Y \mid m^{\alpha} : X \rightarrow R, \alpha \in \Delta\}$ a coordinate system for Y, $\text{Im}(m^{\alpha} | Y) = R_{\alpha} \blacktriangledown \alpha \in \Delta$, $\Pi_{\alpha \in \Delta} R_{\alpha}$ the Cartesian product of the R_{α} and R^{Δ} the collection of all real-valued functions on the index set Δ. The relation $e : Y \rightarrow \Pi_{\alpha \in \Delta} R_{\alpha}$, defined by $e(x) = \gamma_x$, where γ_x is the map $\gamma_x : \Delta \rightarrow \cup_{\alpha \in \Delta} R_{\alpha}$ with $\gamma_x(\alpha) = m^{\alpha}(x)$, is a 1–1 function called the evaluation map induced by M and $\text{Im} e \subset \Pi_{\alpha \in \Delta} R_{\alpha} \subset R^{\Delta}$. If π^{β} is the βth projection

map, then $\pi^\beta \circ e = m^\beta \mid Y$, $\beta \in \Delta$, and $\pi^\beta \circ e(x) = m^\beta(x)$ is the βth coordinate of $x \in Y$. If e' is an evaluation map induced by $\mathcal{M}'$, a coordinate system for Y, then the map $e' \circ e^{-1} : \text{Im} e \xrightarrow{\text{onto}} \text{Im} e'$ is called a coordinate transformation from $\mathcal{M}$ to $\mathcal{M}'$.

Proof: From the definition of e it is obvious that $\text{Im} e \subset \Pi_{\alpha \in \Delta} R_\alpha$. If $R_\alpha = R$ $\blacktriangledown \alpha \in \Delta$ then the Cartesian product is the Cartesian product of copies of R, one for each $\alpha \in \Delta$ and this set is exactly the collection of all real-valued functions on Δ. Hence, $\text{Im} e \subset \Pi_{\alpha \in \Delta} R_\alpha \subset R^\Delta$. Suppose $x = x'$. Then $m^\alpha(x) = m^\alpha(x')$ $\blacktriangledown \alpha \in \Delta$; hence, $\gamma_x = \gamma_{x'}$ and $e(x) = e(x')$ so that e is a function. If $x \neq x'$, then $m^\alpha(x) \neq m^\alpha(x')$ for some $\alpha \in \Delta$ and $\gamma_x \neq \gamma_{x'}$, and therefore $e(x) \neq e(x')$, which establishes that e is $\mathbf{1-1}$. But then e is $1-1$ and onto $\text{Im} e$ and hence its inverse e^{-1} exists and is also $1-1$ and onto. Similarly, e' is $1-1$ and onto $\text{Im} e'$ and this implies that the composite $e' \circ e^{-1}$ is $1-1$ and onto. The only remaining task is to establish $\pi^\beta \circ e = m^\beta \mid Y$. But this follows from the definition of an evaluation map and the definition of a projection map given in Definition I.44. That is, $\pi^\beta \circ e = \gamma_x(\beta)$ $= m^\beta(x)$ $\blacktriangledown x \in Y$.

The next axiom has two purposes. First, it asserts that coordinate systems exist and, second, it makes an assumption about simple processes that brings them into closer agreement with the sketch in Fig. 1.

Axiom I.62: The set of all measurements on X contains a subset $\mathcal{M} = \{m^\alpha : X \to R \mid \alpha \in \Delta\}$ which is a coordinate system for X and a finite subset $\mathcal{M}'$ of cardinality $n \in N$ such that $\mathcal{T} = \{\tau^i = m^i \mid \Sigma \mid m^i \in \mathcal{M}', i = 1, 2, \ldots, n\}$ is a coordinate system for Σ. The elements of $\mathcal{T}$ are called thermodynamic coordinates. For some $\mathcal{M}'$ and some i, say $i = 1$, $\tau^1 = u_{x'}$ for some $x' \in \Sigma$. Further, if $p \in \Pi$ is a simple process and $I = [0, 1] \subset R$ is the unit closed interval, then (1) p is a map $p : I \to X$ and $F(p) = (p(0), p(1))$ and (2) $p \in \mathcal{O}$ iff p is a constant function.

This axiom says that finite-dimensional coordinate systems exist for Σ and that some of these coordinate systems use the internal energy as a coordinate. It does not, however, require all coordinate systems for Σ to have the same dimension nor does it say anything about the dimensions of the coordinate systems for X. The assumptions about simple processes enable us to prove

Theorem I.63: If P is a process of length n_0 on Σ, then P induces a map from the closed-open interval $[0, \infty)$ to X. The induced map, also designated by P, is the function $P : [0, \infty) \to X$ defined by $P(t) = p_n(t - n + 1)$ for $n - 1 \leqslant t \leqslant n$ and $1 \leqslant n \leqslant n_0$ and $P(t) = p_{n_0}(1)$ for $t > n_0$.

Proof: Obvious.

This theorem brings processes into agreement with the sketch of Fig. 2. The following definition merely serves to introduce notation and terminology which will be useful later. Some of the defined quantities are illustrated in Fig. 5.

Definition I.64: Let $\mathcal{M} = \{m^\alpha \mid \alpha \in \Delta\}$ and $\mathcal{T} = \{\tau^i \mid i = 1, 2, \ldots, n\}$ be coordinate systems for X and Σ, respectively, $e(\mathcal{M})$ and $e(\mathcal{T})$ the evaluation maps induced by $\mathcal{M}$ and $\mathcal{T}$, and π^α, $\alpha \in \Delta$ and π^i, $i = 1, 2, \ldots, n$, the corresponding projection maps. Let $m : X \to R$ be a measurement on X, $\tau = m \mid \Sigma$, and write $\tau^i(x) = x^i$ $\blacktriangledown x \in \Sigma$

and $i = 1, 2, \ldots, n$. The function $m \circ e(\mathcal{M})^{-1}$, with domain $\text{Im}[e(\mathcal{M})]$, is called the coordinate representation of m in X. The function $\tau \circ e(\mathcal{T})^{-1}$, with domain $\text{Im}[e(\mathcal{T})]$, is called the coordinate representation of m in Σ. If $P : [0, \infty) \to X$ is a process on Σ, then the collection of functions $\{\pi^\alpha \circ e(\mathcal{M}) \circ P \mid \alpha \in \Delta\}$ is called a coordinate representation of P in X and is usually written as $\{m^\alpha = m_P^\alpha(t) \mid \alpha \in \Delta\}$. If $\text{Im} P \subset \Sigma$, then the collection of functions $\{\pi^i \circ e(\mathcal{T}) \circ P \mid i = 1, 2, \ldots, n\} = \{x^i = x_P^i(t) \mid i = 1, 2, \ldots, n\}$ is called a coordinate representation of P in Σ. If $i_\Sigma : \Sigma \to X$ is the inclusion map, that is, $i_\Sigma(x) = x$ $\blacktriangledown x \in \Sigma$, then the collection of functions $\{\pi^\alpha \circ e(\mathcal{M}) \circ i_\Sigma \circ e(\mathcal{T})^{-1} \mid \alpha \in \Delta\}$ with domain $\text{Im}[e(\mathcal{T})]$ is said to be a coordinate representation of Σ in X and is usually written as $\{m^\alpha = m^\alpha(x^1, x^2, \ldots, x^n) = m^\alpha(x^i) \mid \alpha \in \Delta\}$.

It is not only measurements which have coordinate representations. Suppose $Y \subset X$, Z some set and $\psi : Y \to Z$. Then ψ will have a coordinate representation $\psi \circ (e(\mathcal{M}) \mid Y)^{-1}$. If $Y \subset Z \subset X$, then ψ will also have the representations $\psi \circ (e(\mathcal{T}) \mid Y)^{-1}$. In particular if $Y = \Sigma$, we have the coordinate representation $\psi \circ e(\mathcal{T})^{-1}$ and written as $\psi = \psi(x^i)$. The following two definitions round out the algebraic treatment of thermodynamics. They simply define concepts that appear in other treatments of thermodynamics but which have not been used thus far.

Definition I.65: A process $P : [0, \infty) \to X$ on Σ is said to be quasistatic, or a process in Σ, iff $\text{Im} P \subset \Sigma$.

Definition I.66: Let $\mathcal{T}_1$ and $\mathcal{T}_2$ be coordinate systems for Σ such that $\mathcal{T}_1 - \mathcal{T}_2 = \{u_{x'}\}$ for some $x' \in \Sigma$ and $\mathcal{T}_2 - \mathcal{T}_1 = \{\Theta\}$. The thermodynamic coordinate Θ is said to be an empirical temperature iff $[\Theta(x) \leqslant \Theta(\overline{x})$ iff $u_{x'}(x) \leqslant u_{x'}(\overline{x})]$ and $[\Theta(x) = \Theta(\overline{x})$ iff $u_{x'}(x) = u_{x'}(\overline{x})]$ $\blacktriangledown x$, $\overline{x} \in \Sigma$ such that $\tau(x) = \tau(\overline{x})$ $\blacktriangledown \tau \in \mathcal{T}_1 \cap \mathcal{T}_2$.

The empirical temperature is not unique for if Θ is an
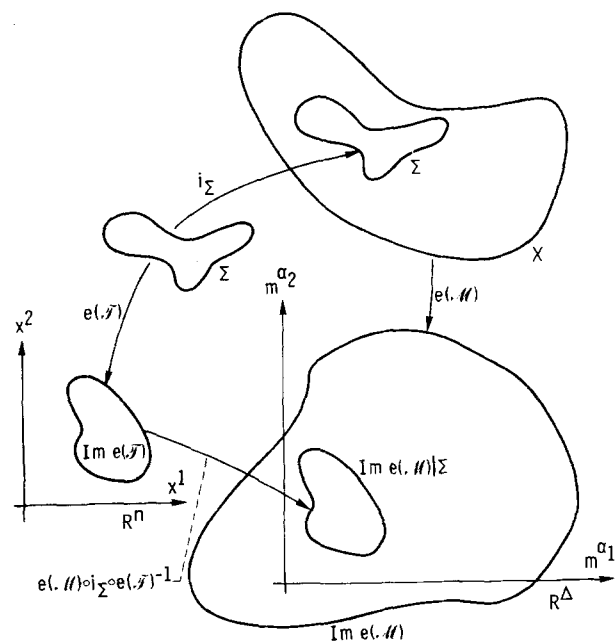
FIG. 5. The relationship between the evaluation map $e(\mathcal{M})$ and the evaluation map $e(\mathcal{T})$.

empirical temperature, then the function Θ_{k_1, k_2} defined by $\Theta_{k_1, k_2}(x) = k_1 \Theta(x) + k_2$ is an equally good temperature for any real numbers $k_1 > 0$, k_2. By Definition I.66 an empirical temperature is regarded merely as an alter ego for the internal energy. This is consistent with the ordinary usage of empirical temperatures in the real world, where one never measures the internal energy directly but instead measures a temperature. One might argue that processes offer a method for measuring the internal energy increments, and hence internal energies, through the first law. But this presupposes that one has previously made unambiguous definitions for both heat and work and this certainly is not the case in thermodynamics, as was pointed out in the Introduction.

The wholly algebraic development of the structure of thermodynamics is now complete, and certainly a few words are in order about what has been accomplished. The fundamental structure of thermodynamics has a completely algebraic character and requires only an independent definition of heat for its implementation and experimental verification. Nowhere in the treatment is there any mention of topology, partitions, or the zeroth law. While work, internal energy, the first law, entropy, quasistatic processes, and temperature are introduced and, in most cases, discussed within the framework of the algebraic theory, they are clearly dispensable concepts. The only indispensable tools are the notions of states, processes connecting these states, and the heat associated with such processes. The absence of topology points to the conclusion that Caratheodory's inaccessaibility axiom is not a basic concept in thermodynamics. The absence of partitions means the algebraic formulation is compatible with partitions and hence seemingly contradicts the statement, made in the Introduction, that a thermodynamics without partitions is effectively a local theory. In a sense this is true, but in a more precise sense there is no contradiction. The reason for this is that the local character of the theory is buried in the assumption that a single point of Σ, rather than a collection of points, represents the state of a real world system. Equivalently, one set of measurements characterizes the real world system.

II. TOPOLOGICAL CONSIDERATIONS

The algebraic prerequisities for the preceding portion of the paper were minimal and as a result that part of the paper is almost completely self-contained. Unfortunately, this is not the situation with respect to the topology and it will be necessary to cite topological theorems that require a somewhat larger background on the part of the reader. All of the necessary material can be found in the recent text on topology written by Willard,[20] although any other comparable text will also suffice. For convenience all citations will be to Willard's book and will be designated by the letter W.

It is obvious that topological considerations become essential when one wishes to speak of continuous functions. It is equally obvious that in a physical theory it is the measurements which one would like to make continuous. But whether or not a given function can be

judged continuous is determined by the topologies assigned to its domain and range. The range for measurements is the set of reals R. In any physical theory one would have to be somewhat addled to consider assigning a topology to R which differs from its usual topology even though other topologies do exist. Henceforth when the topology of R is under discussion it will always be assumed to be the usual topology. What about a topology for Σ? If $(\Sigma, \leqslant_A, =_A)$ were a chain, then it certainly could be endowed with the order topology. However, we have no assurances that this is the case although we do know that Σ is a union of disjoint maximal chains. This does make it possible to choose a subbase for Σ such that the relative topology for a maximal chain would be the order topology. Such a topology, based on order properties, has the disadvantage that it probably would not make measurements on Σ continuous. A topology to be preferred is one which makes measurements on Σ continuous functions. The following scenario is designed to make plausible a choice of topology for the domain. Since the states of Σ are regarded as "simpler" than the states of X, measurements, corresponding to the set $\mathcal{M}'$ of Axiom I.62, are discovered which are adequate for labelling the points of Σ. A topology is assigned to Σ which is just adequate to make the functions $\tau^i \in \mathcal{T}$ continuous. Additional measurements m are invented and deemed acceptable if and only if $m \mid \Sigma$ is continuous in the topology assigned to Σ. Hence, by construction, if $\mathcal{M}$ is a coordinate system for X, then $\mathcal{M}' \subset \mathcal{M}$ and $m \in \mathcal{M} - \mathcal{M}'$ iff $m \mid \Sigma$ is continuous. This story is the basis for the first topological axiom which, for clarity, incorporates a portion of Axiom I.62 in its statement.

Axiom II.1: Let $\mathcal{M} = \{m^\alpha : X \to R \mid \alpha \in \Delta\}$ be a coordinate system for X which contains a finite dimensional subset $\mathcal{M}' \subset \mathcal{M}$ of cardinality $n \in N$ such that (1) $\mathcal{T} = \{\tau^i = m^i \mid \Sigma \mid m^i \in \mathcal{M}', i = 1, 2, \ldots, n\}$ is a coordinate system for Σ and (2) $u_{x'} \in \mathcal{T}$ for some $x' \in \Sigma$. Let $u_{x'} = \tau^1$. Then $m \in \mathcal{M} - \mathcal{M}' \Longrightarrow m \mid \Sigma$ is continuous in the topological space $(\Sigma, \tau(\mathcal{T}))$ where $\tau(\mathcal{T})$ is the weak topology on Σ generated by $\mathcal{T}$.

The definition of a weak topology is given in (W, p. 55, Def. 8.9).

Theorem II.2: Let $(X, \tau(\mathcal{M}))$ be a topological space, where $\tau(\mathcal{M})$ is the weak topology on X generated by $\mathcal{M}$ and $(\Sigma, \tau'(\mathcal{M}))$ be a subspace, that is, $\tau'(\mathcal{M})$ is the relative topology on Σ. Then $\tau'(\mathcal{M}) = \tau(\mathcal{T})$ and hence $(\Sigma, \tau'(\mathcal{M}))$ and $(\Sigma, \tau(\mathcal{T}))$ are homeomorphic.

Proof: By definition of the weak topology (W, p. 55, Def. 8.9) a subbase for $\tau(\mathcal{T})$ is the collection of subsets of Σ given by $\{(m^i \mid \Sigma)^{-1}(U) \mid m^i \in \mathcal{M}'$ and U open in $R\}$. A subbase for $\tau(\mathcal{M})$ is $\{(m^\alpha)^{-1}(U) \mid m^\alpha \in \mathcal{M}$ and U open in $R\}$ and a subbase for $\tau'(\mathcal{M})$ is the collection of sets $\{(m^\alpha \mid \Sigma)^{-1}(U) \mid m^\alpha \in \mathcal{M}$ and U open in $R\}$ (W, p. 59, Exer. 8H.2). Since $\mathcal{M}' \subset \mathcal{M}$, the subbase for $\tau(\mathcal{T})$ is a subset of the subbase for $\tau'(\mathcal{M})$ and therefore $\tau(\mathcal{T}) \subset \tau'(\mathcal{M})$. But, by Axiom II.1, $m^\alpha \mid \Sigma$ is continuous $\forall \alpha \in \Delta$ in $(\Sigma, \tau(\mathcal{T}))$ and, by (W. p. 44, Th. 7.2a, b), $(m^\alpha \mid \Sigma)^{-1}(U)$ is open and hence $(m^\alpha \mid \Sigma)^{-1}(U) \in \tau(\mathcal{T})$ $\forall \alpha \in \Delta$ and then $\tau'(\mathcal{M}) \subset \tau(\mathcal{T})$. Therefore, $\tau'(\mathcal{M}) = \tau(\mathcal{T})$, and the identity map on Σ is a homemorphism.

We have now succeeded in converting Σ and X into topological spaces based on measurements alone. What

can be said about the resulting topologies?

Theorem II.3: Let $\mathcal{M}$ and $\mathcal{T}$ be the coordinate systems of Axiom II.1, R^Δ and R^n Cartesian products of the reals endowed with the product (Tychonoff) topology, and $e(\mathcal{M})$ and $e(\mathcal{T})$ the evaluation maps induced by $\mathcal{M}$ and $\mathcal{T}$, respectively. The space $(\Sigma, \tau(\mathcal{T}))$ is a metrizable space embedded in R^n by $e(\mathcal{T})$. The space $(X, \tau(\mathcal{M}))$ is a Hausdorff space embedded in R^Δ by $e(\mathcal{M})$ and it is a metrizable space if the cardinality of Δ is finite.

Proof: The embedding of $(\Sigma, \tau(\mathcal{T}))$ and $(X, \tau(\mathcal{M}))$ follows immediately from (W, p. 56, Th. 8.12) since $\tau(\mathcal{T})$ and $\tau(\mathcal{M})$ are the respective weak topologies. The space R^n is metric and hence so too is any subspace. Since $(\Sigma, \tau(\mathcal{T}))$ is homeomorphic to a subspace of R^n and metrizability is a topological property (W, p. 49, Exer. 7I.5), it must be a metrizable space. If Δ has finite cardinality, then the previous discussion also applies to $(X, \tau(\mathcal{M}))$. The reals are a metric space and thus Hausdorff (W, p. 86, Example 13.6b). But then R^Δ is a nonempty product space such that each factor is Hausdorff and hence R^Δ is Hausdorff (W, p. 87, Th. 13.8b). Since every subspace of a Hausdorff space is Hausdorff (W, p. 87, Th. 13.8 a) and $(X, \tau(\mathcal{M}))$ is homeomorphic to a subset of R^Δ, it too is Hausdorff.

As a consequence of this theorem we can prove two corollaries. The first one interrelates $\mathrm{Im}[e(\mathcal{T})]$ and $\mathrm{Im}[e(\mathcal{M})]$, while the second one offers a suitable metric for $(\Sigma, \tau(\mathcal{T}))$.

Corollary II.4: The space $\mathrm{Im}[e(\mathcal{T})]$ is embedded in $\mathrm{Im}[e(\mathcal{M})]$ by $e(\mathcal{M}) \circ i_\Sigma \circ e(\mathcal{T})^{-1}$.

Proof: Since $e(\mathcal{T})$ is a homeomorphism of Σ onto $\mathrm{Im}[e(\mathcal{T})]$, then $e(\mathcal{T})^{-1}$ is a homeomorphism of $\mathrm{Im}[e(\mathcal{T})]$ onto Σ. The inclusion map i_Σ is a homeomorphism of Σ into X, and $e(\mathcal{M})$ is a homeomorphism of X onto $\mathrm{Im}[e(\mathcal{M})]$. Hence, the composition $e(\mathcal{M}) \circ i_\Sigma \circ e(\mathcal{T})^{-1}$ is a homeomorphism of $\mathrm{Im}[e(\mathcal{T})]$ into $\mathrm{Im}[e(\mathcal{M})]$.

The embedding covered by Corollary II.4 is illustrated in Fig. 5.

Corollary II.5: The function $\bar{\rho}: \Sigma \times \Sigma \to R$, defined by $\bar{\rho}(x, \bar{x}) = [\sum_{i=1}^n (x^i - \bar{x}^i)^2]^{1/2}$, where $x^i = \tau^i(x)$ and $\tau^i \in \mathcal{T}$, is a metric on Σ and generates the topology $\tau(\mathcal{T})$ on Σ.

Proof: Suppose (Z, τ) is a topological space, (M, ρ) a metric space, and (M, τ_ρ) a topological space with the metric topology τ_ρ induced by ρ. If (Z, τ) is homeomorphic to (M, τ_ρ) and g is the homeomorphism, then it can be shown that $\bar{\rho}: Z \times Z \to R$ defined by $\bar{\rho}(z, \bar{z}) = \rho(g(z), g(\bar{z}))$ is a metric on Z and induces the topology τ. Now $e(\mathcal{T})$ is a homeomorphism onto $\mathrm{Im}[e(\mathcal{T})] \subset R^n$ whose metric is the restriction of the metric on R^n. Hence set Z to Σ, g to $e(\mathcal{T})$, and for ρ use the restriction of the usual metric on R^n.

More details about the topological structure of Σ may be obtained by postulating some additional properties for processess.

Axiom II.6: $\forall x, x' \in \Sigma$ $\exists$ some quasistatic process P such that $x \xrightarrow{P} x'$. Further, if p is a simple process, then p is a continuous function.

Theorem II.7: The topological space $(\Sigma, \tau(\mathcal{T}))$ is pathwise connected and hence connected. If $m|\Sigma$ is not a constant function, then $\mathrm{Im}(m|\Sigma)$ is an interval in R $\forall m \in \mathcal{M}$, where $\mathcal{M}$ is the coordinate system of Axiom II.1.

Proof: Suppose P is a quasistatic process of length n_0 and $x \xrightarrow{P} x'$. Then it follows from Definition I.65 that $\mathrm{Im}p_n \subset \Sigma$ $\forall n \in N$, where $p_n: I \to \Sigma$ is continuous by Axiom II.6. Define $\tilde{p}: I \to \Sigma$ by $\tilde{p}(t) = p_n(n_0 t - n + 1)$ $\forall n \leqslant n_0$ and t such that $(n-1)/n_0 \leqslant t \leqslant n/n_0$. Then $\tilde{p}$ is a function and its restriction to the closed interval $[(n-1)/n_0, n/n_0]$ is continuous because p_n continuous. Further, $\cup_{n=1}^{n_0} [(n-1)/n_0, n/n_0] = [0,1]$ and by (W, p. 48, Exer. 7D.2) or by the repeated application of (W, p. 45, Th. 7.6) it follows that $\tilde{p}$ is continuous. Thus, $\forall x, x' \in \Sigma$ $\exists$ a continuous function $\tilde{p}$ such that $\tilde{p}(0) = x$ and $\tilde{p}(1) = x'$, and by (W, p. 197, Def. 27.1), Σ is pathwise connected, and by (W, p. 197, Th. 27.2), Σ is connected. By Axiom II.1 $m|\Sigma$ is continuous on $(\Sigma, \tau(\mathcal{T}))$ $\forall m \in \mathcal{M}$, for either $m \in \mathcal{M}'$ and then $m|\Sigma$ is continuous by the choice of the topology $\tau(\mathcal{T})$, or $m \in \mathcal{M} - \mathcal{M}'$ and is also continous. But by (W, p. 192, Th. 26.3) $\mathrm{Im}(m|\Sigma)$ is connected since it is the continuous image of a connected space. If $m|\Sigma$ is not a constant function, then $\mathrm{Im}(m|\Sigma) \subset R$ contains at least two points. But the only connected subsets of R with at least two points are the intervals.

To establish Theorem II.7, it was only necessary to postulate that quasistatic simple processes were continuous, whereas Axiom II.6 assumes all simple processes to be continuous. This broader assumption enables us to prove, in the next theorem, that all processes are continuous. Boyling[9] makes the more restrictive assumption that they are continuous and possess continuous derivatives of all orders.

Theorem II.8: If $P: [0, \infty) \to X$ is a process, then P is continuous, as are $e(\mathcal{M}) \circ P$ and $\pi^\alpha \circ e(\mathcal{M}) \circ P$ $\forall \alpha \in \Delta$. If P is quasistatic, then $e(\mathcal{T}) \circ P$ and $\pi^i \circ e(\mathcal{T}) \circ P$, $i = 1, 2, \ldots, n$, are continuous.

Proof: Suppose P is a process of length n_0. Then by the definition of P given in Theorem I.63, the restriction of P to $[n-1, n]$ is continuous since it is p_n $\forall$ $n \leqslant n_0$. Now $\cup_{n=1}^{n_0} [n-1, n] = [0, n_0]$ and by (W, p. 48, Exer. 7D.2) or by the repeated application of (W, p. 45, Th. 7.6) it follows that the restriction of P to $[0, n_0]$ is continuous. If $r \in [n_0, \infty]$, then the restriction of P to $[n_0, r]$ is also continuous since it is a constant function and by (W, p. 45, Th. 7.6), P is continuous on $[0, r]$. Now suppose $r_1 \in [0, \infty)$. Then choose r so that $n_0 \leqslant r$ and $r_1 \leqslant r$ for then $r_1 \in [0, r]$ and thus P is continuous at r_1 $\forall r_1 \in [0, \infty)$ and hence continuous by (W, p. 44, Def. 7.1). The evaluation map $e(\mathcal{M})$ is a homeomorphism and therefore continuous, which means that $e(\mathcal{M}) \circ P$ is continuous since the composition of continuous functions is continuous by (W, p. 45, Th. 7.3). Similarly, the continuity of $m^\alpha = \pi^\alpha \circ e(\mathcal{M})$ assures the continuity of $\pi^\alpha \circ e(\mathcal{M}) \circ P = m^\alpha \circ P$. If P is quasistatic, then $\mathrm{Im}P \subset \Sigma$ and the continuity of $e(\mathcal{T})$ and $\pi^i \circ e(\mathcal{T}) = \tau^i$ make $e(\mathcal{T}) \circ P$ and $\pi^i \circ e(\mathcal{T}) \circ P = \tau^i \circ P$ continuous.

Except for the notion of the reverse of a process, all of the abstract ideas associated with states and pro-

cesses have now been made more concrete by relating them to measurements. The reverse process will be made less abstract by the next axiom, where the usual definition of a reverse process will be given. This should be viewed as a realization of the function f of Axiom I.1.

Axiom II.9: If p is a simple process on Σ, then the reverse of p is given by $p^*(t) = p(1-t)$ $\mathbf{V}$ $t \in [0,1]$.

Theorem II.10: If P is a process of length n_0 on Σ, then the reverse of P is given by $P^*(t) = P(n_0 - t)$ for $t \leq n_0$ and $P^*(t) = P(0)$ for $t > n_0$. P^* is quasistatic iff P is quasistatic.

Proof: We need only consider $0 \leq t \leq n_0$, for which $P(t) = p_n(t - n + 1)$ when $n - 1 \leq t \leq n$ by Theorem I.63. But if $n - 1 \leq t \leq n$, then $n_0 - n \leq n_0 - t \leq n_0 - n + 1$ and $0 \leq n_0 - t - (n_0 - n) \leq 1$ or $0 \leq n - t \leq 1$ and, hence, $P(n_0 - t) = p_{n_0 - n + 1}(n - t)$ when $n - 1 \leq t \leq n$. But, using Axiom II.9, we obtain $P(n_0 - t) = p^*_{n_0 - n + 1}(1 - n + t)$ when $n - 1 \leq t \leq n$ and by Theorem I.63 this implies $P(n_0 - t) = P^*(t)$ for $t \leq n_0$, where Definition I.14 was also used. This result now shows that $\mathrm{Im}P = \mathrm{Im}P^*$ and hence P^* is quasistatic iff P is quasistatic.

In view of this theorem and Axiom II.9 a process is to be considered reversible if and only if a system can be made to pass through the same states but in reverse order. Hence, this rules out the situation depicted in Fig. 1 where a process and its reverse do not coincide. We now have available not only algebraic structure but also topological structure on Σ. The next two theorems combine the two structures to obtain results which cannot be obtained from either alone. The first theorem concerns itself with empirical temperatures, while the second one deals with empirical entropies. In these theorems and for the balance of the paper the subscript on the internal energy will be suppressed and the internal energy will simply be denoted by u rather than u_x.

Theorem II.11: Let Θ be an empirical temperature and denote its coordinate representation in Σ by the symbol θ, that is, $\theta = \theta(x^1, x^2, \ldots, x^n) = \theta(u, x^2, \ldots, x^n)$. Then θ is a monotone increasing function of the internal energy. Further, if θ is differentiable with respect to u in $Y \subset \Sigma$, then $\partial \theta / \partial u > 0$ in Y. If $\partial \theta / \partial u$ is continuous in Y, then $\mathbf{E}$ a unique solution $u = u(\theta, x^2, \ldots, x^n)$ in Y such that u and $\partial u / \partial \theta \equiv c$ are continuous with respect to θ. The heat capacity c satisfies the condition $c > 0$ in Y.

Proof: Since Θ is a temperature, it is the restriction of a measurement to Σ and hence is continuous and has the coordinate representation $\theta = \theta(u, x^2, \ldots, x^n)$. But then Definition I.66 is just the requirement that θ be monotone increasing. If it is differentiable, then the monotone increasing property shows that $\partial \theta / \partial u > 0$. The existence and continuity of $u(\theta, x^2, \ldots, x^n)$ and $\partial u / \partial \theta$ follow from a standard theorem known as the implicit function theorem. Further since $(\partial u / \partial \theta)(\partial \theta / \partial u) = 1$, we have $c > 0$.

Theorem II.12: Let φ be an empirical entropy for some chain in Σ and $\phi = \phi(u, x^2, \ldots, x^n)$ its coordinate

representation. Suppose ϕ is a monotone increasing function of u, continuous and possessing continuous first partial derivatives with respect to x^i, $i = 1, 2, \ldots, n$, in some subchain C; then, in C, $T_\phi d\phi = du - [(\partial u / \partial x^2)_\phi dx^2 + (\partial u / \partial x^3)_\phi dx^3 + \cdots + (\partial u / \partial x^n)_\phi dx^n]$, where $1/T_\phi = \partial \phi / \partial u > 0$. The function T_ϕ is called the absolute temperature in C associated with the empirical entropy ϕ.

Proof: From the conditions in the statement of the theorem it follows that $\partial \phi / \partial u > 0$. Further, $d\phi = (\partial \phi / \partial u) du + (\partial \phi / \partial x^2) dx^2 + \cdots + (\partial \phi / \partial x^n) dx^n$. Since $\partial \phi / \partial u \neq 0$, we can divide by $\partial \phi / \partial u$ and use $(\partial \phi / \partial x^i)/(\partial \phi / \partial u) = -(\partial u / \partial x^i)_\phi$, $i = 2, \ldots, n$.

Although T_ϕ has been called a temperature, it might not be an empirical temperature for two reasons. First, it might not be a monotone increasing function of u and, second, it might not correspond to any measurement which could be performed on X. If T_ϕ were differentiable, then it would be a monotone increasing function of u if and only if $\partial^2 \phi / \partial u^2 < 0$. The fact that T_ϕ should be coupled to a particular empirical entropy is obvious if one considers the entropy $k\phi$, where k is a positive real number. For then it follows that $kT_{k\phi} = T_\phi$.

The differential form which appears in Theorem II.12 is very much like the form of the first law displayed in thermodynamics textbooks. It cannot, however, be the first law because it is defined only for increments in the chain $C \subset \Sigma$, while the first law must be defined for any process P which need not even be quasistatic. I wish to pursue the consequences of assuming a differentiable form for the first law but then processes themselves must be assumed differentiable in the sense that the functions $\pi^\alpha \circ e(\mathcal{M}) \circ P = m^\alpha \circ P$, which are the coordinate representation of P, are differentiable. These items are incorporated into the next axiom.

Axiom II.13: All processes P are differentiable, that is, $\pi^\alpha \circ e(\mathcal{M}) \circ P$ is differentiable $\mathbf{V}$ $\alpha \in \Delta$. Let $\mathcal{M}, \mathcal{M}'$, and $\mathcal{T}$ be the sets of Axiom II.1 and define a finite subset of the index set Δ by $\{\alpha^i \mid m^{\alpha^i} \in \mathcal{M}'$ and $m^{\alpha^i} \mid \Sigma = \tau^i \in \mathcal{T}\}$. Then $\mathbf{E}$ n real-valued functions $\beta_i : X \to R$, with coordinate representations $B_i = \beta_i \circ e(\mathcal{M}) = B_i(m^\alpha)$, $i = 1, 2, \ldots, n$, in X and $B_1 = 1$. The heat increment is the differential form $d\!Q = B_1 dm^{\alpha_1} + \cdots + B_n dm^{\alpha_n}$ and the work increment is $dW = -[B_2 dm^{\alpha_2} + \cdots + B_n dm^{\alpha_n}]$. For any process P the heat and work are $Q(P) = \int_0^\infty dt\, dQ(P)/dt$ and $W(P) = \int_0^\infty dt\, dW(P)/dt$, where $d\!Q(P)/dt = B_1 dm_P^{\alpha_1}/dt + \cdots + B_n dm_P^{\alpha_n}/dt$ and $d\!W(P)/dt = -[B_2 dm_P^{\alpha_2}/dt + \cdots + B_n dm_P^{\alpha_n}/dt]$. If P is quasistatic, then the notation $d\!Q(P)/dt = d\!q(P)/dt \equiv A_1 dx_P^1/dt + \cdots + A_n dx_P^n/dt$ and $d\!W(P)/dt = dw(P)/dt \equiv -[A_2 dx_P^2/dt + \cdots + A_n dx_P^n/dt]$ will be used, where $A_i(x^j) = B_i[m^\alpha(x^j)] \neq 0$ $\mathbf{V}$ $x \in \Sigma$.

The definitions contained in Axiom II.13 lead to the result $Q(P) + W(P) = \int_0^\infty dt\, dm_P^1/dt = m^1(\infty) - m^1(0) \equiv \Delta u$ since $m^1 \mid \Sigma = u$ and the end points of P are in Σ, and also to the result $q(P) + w(P) = \int_0^\infty dt\, dx_P^1/dt = \Delta u$. Thus, Axiom II.13 contains the first law for any process as well as its restriction to quasistatic processes. Two things should be pointed out in connection with Axiom II.13. First, the functions β_i were not required to be measurements, although they could be. Second, while

the quasistatic form of the first law is obtained from the general form, it will not be possible to reverse the procedure if Σ is a proper subset of X. As has already been pointed out, the notion of entropy is not dependent upon the first law and may be discussed purely in terms of heat. But the usual treatments of thermodynamics do associate an entropy with the quasistatic form of the first law. This connection can be investigated provided that some additional assumptions are made. These assumptions will be made here since they lead to results of a topological nature. However, the connection between the quasistatic form of the first law and the entropy will be treated in the following section.

Axiom II.14: Let C be a maximal chain in $(\Sigma, \leq_A, =_A)$. Then (1) $x =_A x'$ iff $\exists$ a quasistatic process P such that $x \xrightarrow{P} x'$ and $\mathit{dq}(P)/dt = 0 \ \forall \ t \in [0, \infty)$, (2) $\forall x$, $x' \in C \ \exists$ some process P such that $x \xrightarrow{P} x'$ and Im $P \subset C$, and (3) every continuous differentiable map $P: [0, \infty) \to \Sigma$. such that $P|[r, \infty)$ is a constant function for some $r < \infty$, is a process.

Theorem II.15: Let C be a maximal chain in $(\Sigma, \leq_A, =_A)$ and $x \in C$. Then C and $[x]$ are pathwise connected and hence connected.

Proof: The connectivity and pathwise connectivity of C follow from Axiom II.14(2) exactly as the corresponding properties for Σ followed from Axiom II.6 in Theorem II.7. To establish the result for $[x]$, we observe that by definition $[x] = \{x' | x' \in \Sigma \text{ and } x' =_A x\}$. But $x' =_A x \Rightarrow x =_A x'$ and by Axiom II.14(1) $\exists$ a quasistatic process P such that $x \xrightarrow{P} x'$ and $dq(P)/dt = 0 \ \forall \ t \in [0, \infty)$. Let $r_1 \in [0, \infty)$ and define $\tilde{P}:[0, \infty) \to \Sigma$ by $\tilde{P}|[0, r_1]$ $= P|[0, r_1]$ and $\tilde{P}(t > r_1) = P(r_1)$. Then by Axiom II.14(3) $\tilde{P}$ is a quasistatic process and by construction $\mathit{dq}(\tilde{P})/dt = 0 \ \forall \ t \in [0, \infty)$; hence, $x =_A \tilde{P}(r_1) \ \forall \ r_1 \in [0, \infty)$ and thus Im$P \in [x]$. But in the manner of Theorem II.7, P induces a map from the unit interval I to Σ whose range then lies in $[x]$. Now suppose x', $x'' \in [x]$; then $x' =_A x =_A x''$ and obviously $\exists$ a path from x' to x'' in $[x]$. Thus, $[x]$ is pathwise connected and hence connected.

Corollary II.16: Let $\mathcal{M}$ be the coordinate system of Axiom II.1 and $m \in \mathcal{M}$. If $m|C(m|[x]|)$ is not a constant function, then Im$m|C(\text{Im}m|[x])$ is an interval in R.

Proof: See the proof of the corresponding property in Theorem II.7 using the fact that the restriction of a continuous function is continuous (W, p. 45, Th. 7.5).

Observe that this section and the previous one have reduced the abstract ideas of states, processes, coordinate systems and topological structure to the concrete experimental idea of measurement. Yet in spite of their commonality, there exists a strange gulf that separates the two sections. It is significant that, while quasistatic processes appear quite prominently in this section, they are essentially nonexistent in the preceding section. The converse is true of reversible processes. A second significant difference between the two sections is that in the previous section it was necessary to distinguish between physical and nonphysical processes, whereas such a distinction is not even mentioned here. It is not even necessary to regard a quasistatic process as a physical process and hence it need not be possible to carry out such a process in the laboratory. Based solely on axiom II.14(1), it is tempting to consider the condition

$\mathit{dq}(P)/dt = 0 \ \forall \ t \in [0, \infty)$ as the definition of a reversible adiabatic process. Although it is done conventionally, it is neither necessary nor desirable for it imposes additional constraints on the theory. More properly Axiom II.14(1) should be viewed as an alternative way of determining adiabatic equality just as was done with Axiom I.54. The counterpart of Theorem I.55 is

Theorem II.17: $x <_A x'$ iff $x \leq_A x'$ and $\not\exists$ a quasistatic process P such that $(x \xrightarrow{P} x'$ and $\mathit{dq}(P)/dt = 0 \ \forall \ t \in [0, \infty))$.

Proof: From Theom I.21 $x <_A x'$ iff $x \leq_A x'$ and $x \neq_A x'$. Thus, the theorem follows from the contraposition of Axiom II.14(1).

III. INTEGRABILITY CONDITIONS

Considerations about integrability in thermodynamics specifically focus on the relationships among (1) the entropy, if it exists, (2) its total differential in the form given in Theorem II.12, if it exists, and (3) the quasistatic differential form of the first law, dq, if it exists. This topic itself would not exist were it not for the existence of Axiom II.14. Its counterpart, Axiom I.54, is supported by a substantial accumulation of experimental fact but it is not obvious that Axiom II.14 is valid. Its verification requires one to know the quasistatic differential form of the first law, that is, the $A_i = A_i(x^j)$, $i = 1, 2, \ldots, n$, and either to construct mathematically the quasistatic processes such that $\mathit{dq}(P)/dt = 0 \ \forall \ t \in [0, \infty)$ or else to carry out such processes experimentally. But these processes are quasistatic and the (unproven) consensus seems to be that quasistatic processes are not physically realizable processes (Wilson,[3] p. 9; Landsberg,[6] 1961, p. 35; and Callen,[11] p. 60). This, if true, leaves us with only the possibility of a mathematical test. The mathematical test is generally based upon a Pfaffian form, the quasistatic differential form of the first law, coupled with Caratheodory's theorem and Caratheodory's adiabatic inaccessibility axiom. Caratheodory's theorem provides necessary and sufficient conditions for the integrability of a Pfaffian form in terms of the properties of its integral curves. Landsberg[6] (1961, p. 50) gives a proof of the theorem based on reduction of the Pfaffian to canonical form, as does Bernstein,[8] while Buchdahl[7] [1949 (p. 44, p. 212), 1954, 1955] offers proofs in terms of integral curves. The use of Caratheodory's theorem is an indirect approach to the integrability problem since it uses integral curves rather than the Pfaffian itself. This oblique attack on the integrability problem might have an advantage if it eliminated the need to know the functions $A_i(x^j)$, as would be the case if quasistatic processes were physical processes. It would also have an advantage if it eliminated some restrictions on the functions $A_i(x^j)$. Neither of these criteria are satisfied by the approach based on Caratheodory's theorem. Further, this method has not been wholeheartedly accepted by thermodynamicists.

Of course, a direct frontal assault on the Pfaffian dq should also be capable of dealing with questions of integrability. The basis for such an approach is already partially available in the literature of differential equations.[21] The analysis of this section is assumed to take place on some open connected subset, say $V \subset \text{Im}e(\mathcal{T})$

$\subset R^n$, and it will be useful to define some standard notation and terminology.

Definition III.1: Let $V \subset \mathrm{Im} e(\mathcal{T}) \subset R^n$ be an open connected subset of $\mathrm{Im} e(\mathcal{T})$. (1) A real-valued function $\psi: V \to R$ is said to be of class $C^{(k)}$, k a nonnegative integer, iff all partial derivatives up to and including those of order k exist and are continuous. (2) A collection of real-valued functions $\{\psi_i : V \to R \mid i = 1, 2, \ldots, p\}$ is said to be functionally dependent in V iff $\exists$ a relation $\Psi(\psi_1, \psi_2, \ldots, \psi_p) = 0$ which does not explicitly depend upon the coordinates x^k, $k = 1, 2, \ldots, n$. If the functions are not dependent, they are said to be independent. (3) If $\exists$ a collection $\{z^i = z^i(x^j) \mid i = 1, 2, \ldots, n\}$ of independent, class $C^{(1)}$, functions on V such that for some nonnegative integer $\bar{n} \leqslant n$, $z^i = c^i$, $i > \bar{n}$, where c^i is a constant, then V is said to be an $\bar{n}$-dimensional subspace of $e(\mathcal{T})$ and $z^1, z^2, \ldots, z^{\bar{n}}$ are called intrinsic coordinates for V. The functions $x^i = x^i(z^1, \ldots, z^{\bar{n}}, c^{\bar{n}+1}, \ldots, c^n) = x^i(z^1, \ldots, z^{\bar{n}})$ are said to give a parametric representation of V. (4) When an index is repeated in a term, *once as a subscript and once as a superscript*, summation over the range of that index is understood.

The existence of the parametric representation of V in Definition III.1(3) is a consequence of the implicit function theorem which also shows the representation to be of class $C^{(1)}$. The next theorem and its corollary are also standard consequences of the implicit function theorem and are cited without proof. They will be used in the subsequent analysis.

Theorem III.2: Consider a collection $\{\psi_1, \psi_2, \ldots, \psi_p\}$ of p real-valued, class $C^{(1)}$ functions on V, where p is a positive integer. Then $\exists$ $p - r$ relations connecting these functions iff rank $J = r$, where J is the Jacobian matrix $\partial(\psi_1, \psi_2, \ldots, \psi_p)/\partial(x^1, x^2, \ldots, x^n)$ and rank J is its rank.

Corollary III.3: Two real-valued functions of class $C^{(1)}$ on V, ψ_1 and ψ_2, are functionally dependent in V iff $\partial(\psi_1, \psi_2)/\partial(x^i, x^j) = 0$ $\forall$ i, j.

The next theorem and its corollary make a partial connection between an empirical entropy and the quasistatic differential form of the first law.

Theorem III.4: Let φ be an empirical entropy for a maximal chain in $(\Sigma, \leqslant_A, =_A)$ and $\phi = \phi(x^i)$ its coordinate representation. Suppose $\exists$ an open, connected subchain C such that $\varphi \mid C$ is not a constant function, C is an $\bar{n}$-dimensional subspace of $e(\mathcal{T})$ with parametric representation $x^i = x^i(z^1, z^2, \ldots, z^{\bar{n}})$, and, in C, ϕ is of class $C^{(1)}$. If P is a quasistatic process such that $\mathrm{Im} P \subset C$, then $đq(P) = 0 \Rightarrow \mathrm{Im} P \subset [x]$ for some $x \in C$ iff $P \circ \phi = \mathrm{const}$ iff $d(P \circ \phi) = 0$.

Proof: $P \circ \phi = \mathrm{const}$ iff $d(P \circ \phi) = 0$ is obvious. Now $đq(P) = 0 \Rightarrow đq(P)/dt = 0$ $\forall t \in [0, \infty)$ and hence $P(r_1) =_A P(r_2)$ $\forall$ $r_1, r_2 \in [0, \infty)$ as in Theorem II.15. If x is any point in $\mathrm{Im} P$, then $\mathrm{Im} P \subset [x]$. But $\forall$ $x' \in [x]$, $\varphi(x') = \varphi(x)$ since φ is an empirical entropy; hence $\phi(x') = \phi(x)$ and thus $P \circ \phi = \mathrm{const}$. Conversely, $P \circ \phi = \mathrm{const}$ $\Rightarrow \varphi(P(r_1)) = \varphi(P(r_2)) \Rightarrow P(r_1) =_A P(r_2)$ $\forall$ $r_1, r_2 \in [0, \infty)$ $\Rightarrow \mathrm{Im} P \subset [x]$ for some $x \in \mathrm{Im} P$.

Corollary III.5: Suppose the conditions of Theorem

III.4 apply and $\mathrm{Im} P \subset [x] \Rightarrow đq(P) = 0$. Then, in C, $đq(P) = 0$ iff $d(P \circ \phi) = 0$ and $\exists$ a real-valued function $\lambda = \lambda(z^1, z^2, \ldots, z^{\bar{n}}) \neq 0$ such that $d\phi = \lambda đq$ and $\partial\phi/\partial z^\alpha = \lambda A_i \partial x^i/\partial z^\alpha$, $\alpha = 1, 2, \ldots, \bar{n}$. If P is a quasistatic process, $x \xrightarrow{P} \bar{x}$ and $\mathrm{Im} P \subset C$, then $[\varphi(\bar{x}) - \varphi(x)]/|\lambda|_{max} \leqslant q(P) \mathrm{sgn}(\lambda) \leqslant [\varphi(\bar{x}) - \varphi(x)]/|\lambda|_{min}$, where $|\lambda|_{max}$, $|\lambda|_{min}$ are the largest and smallest values of the magnitude of λ along P and $\mathrm{sgn}(\lambda)$ is the sign of λ. If $\bar{n} = n$, then $\lambda = \partial\phi/\partial u$, $\partial\phi/\partial x^i = \lambda A_i$, $i = 1, 2, \ldots, n$, and $A_i = -(\partial u/\partial x^i)_\phi$, $i = 2, 3, \ldots, n$.

Proof: Combining the condition in this corollary with $đq(P) = 0 \Rightarrow \mathrm{Im} P \subset [x]$ from the previous theorem establishes the first part of the corollary. By the chain rule for differentiation the condition $đq(P) = 0$ iff $d(P \circ \phi) = 0$ gives $(\partial\phi/\partial z^\alpha) dz^\alpha = (\partial\phi/\partial x^i)(\partial x^i/\partial z^\alpha) dz^\alpha = 0$ iff $A_i(\partial x^i/\partial z^\alpha) \partial z^\alpha = 0$, where α runs from 1 to $\bar{n}$. For the case $\bar{n} = 1$ we have, upon suppressing index α, $d\phi/dz = A_i dx^i/dz$ and take $\lambda = 1$. If $\bar{n} > 1$, then let $\alpha_1, \alpha_2 \in [1, 2, \ldots, \bar{n}]$; now since the z^α are independent, we can choose displacements so that $dz^\alpha = 0$ for $\alpha \neq \alpha_1, \alpha_2$. Then $d\phi = (\partial\phi/\partial z^{\alpha_1}) dz^{\alpha_1} + (\partial\phi/\partial z^{\alpha_2}) dz^{\alpha_2}$ and $đq = (A_i \partial x^i/\partial z^{\alpha_1}) dz^{\alpha_1} + (A_i \partial x^i/\partial z^{\alpha_2}) dz^{\alpha_2}$. Then $đq = 0$ iff $d\phi = 0$ leads to a pair of homogeneous equations for $dz^{\alpha_1}, dz^{\alpha_2}$ which have a nontrivial solution iff the determinant $(\partial\phi/\partial z^{\alpha_1})(A_i \partial x^i/\partial z^{\alpha_2}) - (\partial\phi/\partial z^{\alpha_2})(A_i \partial x^i/\partial z^{\alpha_1})$ vanishes. Varying α_1, α_2, we get

$$\frac{\partial\phi/\partial z^1}{A_i \partial x^i/\partial z^1} = \frac{\partial\phi/\partial z^2}{A_i \partial x^i/\partial z^2} = \frac{\partial\phi/\partial z^{\bar{n}}}{A_i \partial x^i/\partial z^{\bar{n}}} \equiv \lambda.$$

Observe that $\lambda = 0 \Rightarrow \partial\phi/\partial z^\alpha = 0 \Rightarrow \phi = \mathrm{const}$, a contradiction and hence $\lambda \neq 0$. It immediately follows that $d\phi = \lambda đq$. Now $q(P) \mathrm{sgn}(\lambda) = \mathrm{sgn}(\lambda) \int_x^{\bar{x}} d\phi/\lambda = \int_x^{\bar{x}} d\phi/|\lambda|$, where the integrals are line integrals along P from x to $\bar{x}$. Further, we have $[\phi(\bar{x}^i) - \phi(x^i)]/|\lambda|_{max} \leqslant \int_x^{\bar{x}} d\phi/|\lambda| \leqslant [\phi(\bar{x}^i) - \phi(x^i)]/|\lambda|_{min}$, where $\bar{x}^i = r^i(\bar{x})$. Since $\varphi(\bar{x}) = \phi(\bar{x}^i)$, we have then $[\varphi(\bar{x}) - \varphi(x)]/|\lambda|_{max} \leqslant q(P) \mathrm{sgn}(\lambda) \leqslant [\varphi(\bar{x}) - \varphi(x)]/|\lambda|_{min}$. If $\bar{n} = n$, the matrix of partial derivatives is nonsingular and $(\partial x^i/\partial z^j)(\partial z^j/\partial x^k) = \delta_k^i$, where δ_k^i is the Kronecker delta. Thus multiplying $\partial\phi/\partial z^j = \lambda A_i \partial x^i/\partial z^j$ by $\partial z^j/\partial x^k$ gives $\partial\phi/\partial x^k = \lambda A_k$. But with $k = 1$ we have $\partial\phi/\partial x^1 = \partial\phi/\partial u = \lambda A_1 = \lambda$ since $A_1 = 1$. But then $A_k = (\partial\phi/\partial x^k)/(\partial\phi/\partial u) = -(\partial u/\partial x^i)_\phi$ for $i = 2, 3, \ldots, n$.

Corollary III.6: Suppose the conditions of Theorem III.4 and Corollary III.5 are satisfied with $\bar{n}$ not necessarily equal to n. Then (1) $x =_A \bar{x}$ iff $q(P) = 0$ and (2) $x \leqslant_A \bar{x}$ iff $q(P) \mathrm{sgn}(\lambda) \geqslant 0$ $\forall$ quasistatic processes P such that $\mathrm{Im} P \subset C$ and $x \xrightarrow{P} \bar{x}$.

Proof: Suppose $q(P) = 0$. Then from Corollary III.5 $[\varphi(\bar{x}) - \varphi(x)]/|\lambda|_{max} \leqslant 0 \leqslant [\varphi(\bar{x}) - \varphi(x)]/|\lambda|_{min}$. Since $|\lambda|_{max} > 0$ and $|\lambda|_{min} > 0$, we find $\varphi(\bar{x}) - \varphi(x) \leqslant 0$ and $\varphi(\bar{x}) - \varphi(x) \geqslant 0$ and hence, since φ is an entropy, $\varphi(x) = \varphi(\bar{x})$ iff $x =_A \bar{x}$. Conversely, $x =_A \bar{x}$ iff $\varphi(x) = \varphi(\bar{x})$ and therefore $0 \leqslant q(P) \mathrm{sgn}(\lambda) \leqslant 0$. Since $\mathrm{sgn}(\lambda) = \pm 1$, then $q(P) = 0$. If $x \leqslant_A \bar{x}$, then $\varphi(\bar{x}) - \varphi(x) \geqslant 0$ and $q(P) \mathrm{sgn}(\lambda) \geqslant 0$. Conversely if $q(P) \mathrm{sgn}(\lambda) \geqslant 0$, then $[\varphi(\bar{x}) - \varphi(x)]/|\lambda|_{min} \geqslant 0$ and it follows that $\varphi(x) \leqslant \varphi(\bar{x})$ iff $x \leqslant_A \bar{x}$.

This result can be regarded as the quasistatic analog of Axiom I.54 and Theorem I.55. However, it should be kept in mind that some fairly strong assumptions were used to derive it. Further, it may not be valid for a maximal chain but only for some subchain.

Now that we have treated one aspect of the connection between entropy and the quasistatic first law, we should also consider the converse problem. In the direct problem we have demonstrated that if an entropy φ is assumed to exist, then under suitable conditions there exists a function λ such that the coordinate representation ϕ of the entropy satisfies $d\phi = \lambda\,\vec{d}q$. In the converse problem we do not assume the existence of an entropy but ask under what conditions does there exist a function, analogous to λ, which converts $\vec{d}q$ into the total differential of some function and under what additional conditions does this latter function become an entropy. We first consider the integrability of $\vec{d}q$ and subsequently examine when the resulting function becomes an entropy. As mentioned before, Caratheodory's theorem is one method of dealing with the integrability problem. I shall now discuss a more direct approach and begin with a definition of terminology.

Definition III.7: Let V be an open connected subset of $e(\mathcal{T})$. The differential form $\vec{d}q = A_k\,dx^k$ is said to be integrable on V iff $\exists$ functions $\phi = \phi(x^i)$ and $M = M(x^i) \neq 0$ such that, on V, $d\phi = M\,\vec{d}q = MA_k\,dx^k$. The function ϕ is called a pseudopotential and the function M is called an integrating factor.

The next theorem, although well known, is cited proven because it will be used in the discussion of the integrability problem and because its method of proof is similar to the proof used in the integrability problem (Theorem III.9).

Theorem III.8: Let V be an open connected subset of $\mathrm{Im}\,e(\mathcal{T})$ and $\{\psi_i \,|\, i = 1, 2, \ldots, n\}$ a collection of $C^{(1)}$ functions on V. The system of partial differential equations $\partial\phi/\partial x^i = \psi_i$ possesses a solution on V iff the conditions $\partial\psi_i/\partial x^k - \partial\psi_k/\partial x^i = 0$, $i, k = 1, 2, \ldots, n$, are satisfied.

Proof: Suppose ϕ is a solution. Then $\partial^2\phi/\partial x^k\,\partial x^i = \partial^2\phi/\partial x^i\,\partial x^k \Rightarrow \partial\psi_i/\partial x^k - \partial\psi_k/\partial x^i = 0$. Obviously, if $n = 1$, the ordinary differential equation possesses a solution given by the indefinite integral $\phi = \int \psi_1\,dx^1$. Now proceeding by induction, assume the theorem true for $n-1$, use Greek indices $\alpha, \beta, \cdots = 1, 2, \ldots, n-1$, and suppose ϕ satisfies $\partial\phi/\partial x^\alpha = \psi_\alpha$, where $\phi = \phi(x^\alpha; x^n)$ and x^n is regarded as a parameter. Let $\Delta \equiv \psi_n - \partial\phi/\partial x^n$. Then $\partial\Delta/\partial x^\alpha = \partial\psi_n/\partial x^\alpha - \partial^2\phi/\partial x^n\,\partial x^\alpha = \partial\psi_n/\partial x^\alpha - \partial\psi_\alpha/\partial x^n = 0$ and therefore $\Delta = \Delta(x^n)$. Let $\Phi = \phi + \int \Delta(x^n)\,dx^n$. Then $\partial\Phi/\partial x^\alpha = \partial\phi/\partial x^\alpha = \psi_\alpha$ and $\partial\Phi/\partial x^n = \partial\phi/\partial x^n + \Delta = \psi_n$, and thus Φ is a solution of $\partial\phi/\partial x^i = \psi_i$, $i = 1, 2, \ldots, n$.

The introductory remarks to this section contained an allusion to the availability of results in the literature which would provide a basis for a direct approach to the integrability of $\vec{d}q$. Specifically, I had in mind the contents of the next theorem, which are available in Chap. 1 of Forsyth.[21] Instead of merely quoting the results, the theorem will be proved here because of its importance, its simplicity of proof, and because the method of proof differs from that found in Forsyth and may be more appealing to some.

Theorem III.9: Let V be an open connected subset of $\mathrm{Im}\,e(\mathcal{T})$, $\vec{d}q = A_k\,dx^k$, where the functions $A_k = A_k(x^i) \neq 0$, $k = 1, 2, \ldots, n$, are of class $C^{(1)}$ on V. Then $\vec{d}q$ is integrable on V iff $A_i(\partial A_k/\partial x^j - \partial A_j/\partial x^k) + A_j(\partial A_i/\partial x^k - \partial A_k/\partial x^i) + A_k(\partial A_j/\partial x^i - \partial A_i/\partial x^j) = 0$ $\forall$ $i, j, k = 1, 2, \ldots,$

n. Further, an integrating factor is not unique, and any function $M = M(x^i) \neq 0$ is an integrating factor iff it is a solution of the system of first order partial differential equations $A_i\,\partial M/\partial x^j - A_j\,\partial M/\partial x^i = M(\partial A_j/\partial x^i - \partial A_i/\partial x^j)$, $i, j = 1, 2, \ldots, n$.

Proof: Suppose $\vec{d}q$ is integrable. Then $M\,\vec{d}q = MA_k\,dx^k = (\partial\phi/\partial x^k)dx^k$ or $\partial\phi/\partial x^k = MA_k$. Using $\partial^2\phi/\partial x^i\,\partial x^j = \partial^2\phi/\partial x^j\,\partial x^i$ gives $\partial(MA_j)/\partial x^i = \partial(MA_i)/\partial x^j$ and performing the differentiations gives the differential equations $A_i\,\partial M/\partial x^j - A_j\,\partial M/\partial x^i = M(\partial A_j/\partial x^i - \partial A_i/\partial x^j)$. Multiplying this by A_k gives $MA_k(\partial A_j/\partial x^i - \partial A_i/\partial x^j) = A_kA_i\,\partial M/\partial x^j - A_kA_j\,\partial M/\partial x^i$. A cyclic permutation of indices $k \to i \to j \to k$ followed by addition of the resulting three equations produces $M[A_k(\partial A_j/\partial x^i - \partial A_i/\partial x^j) + A_i(\partial A_k/\partial x^j - \partial A_j/\partial x^k) + A_j(\partial A_i/\partial x^k - \partial A_k/\partial x^i)] = 0$. But since $M(x^i) \neq 0$, the bracketed factor must vanish. Conversely, suppose M is a solution of the system $A_i\,\partial M/\partial x^j - A_j\,\partial M/\partial x^i = M(\partial A_j/\partial x^i - \partial A_i/\partial x^j)$. Then this $\Rightarrow \partial(MA_j)/\partial x^i - \partial(MA_i)/\partial x^j = 0$; by the previous theorem the system $\partial\phi/\partial x^i = MA_i$ possesses a solution ϕ, and, of course, $d\phi = (\partial\phi/\partial x^i)\,dx^i = MA_i\,dx^i = M\,\vec{d}q$. The proof of this theorem is complete if we can establish that the system of partial differential equations for M has a solution when the integrability conditions are satisfied. We must first show the system of equations is consistent; that is, if we select any n of the equations which can be solved for the derivatives $\partial M/\partial x^i$, then these solutions will identically satisfy the remaining equations. Equivalently, we can show that any pair of equations with a common index will imply another equation of the system. Thus, consider the (j, k) and (k, i) members of the system; multiply the former by A_i, the latter by A_j and add the resulting equations to get

$$A_k\left(A_j\frac{\partial M}{\partial x^i} - A_i\frac{\partial M}{\partial x^j}\right)$$
$$= M\left[A_i\left(\frac{\partial A_k}{\partial x^j} - \frac{\partial A_j}{\partial x^k}\right) + A_j\left(\frac{\partial A_i}{\partial x^k} - \frac{\partial A_k}{\partial x^i}\right)\right]$$
$$= -MA_k\left(\frac{\partial A_j}{\partial x^i} - \frac{\partial A_i}{\partial x^j}\right).$$

The last equality is a result of using the integrability conditions. But since $A_k \neq 0$, this implies $A_j\,\partial M/\partial x^i - A_i\,\partial M/\partial x^j = M(\partial A_i/\partial x^j - \partial A_j/\partial x^i)$, which is the (j, i) member of the system. Thus, the system is consistent by virtue of the integrability conditions. Now we complete the proof by induction to demonstrate that a solution exists. For $n = 1$ we have $\vec{d}q = A_1\,dx^1$. If we set $\phi = \int A_1\,dx^1$, then $d\phi = \vec{d}q$ with $M = 1$ and an integrating factor exists for $n = 1$. We now assume that the case $n-1$ is integrable, that is, if Greek indices are used for the range $1, 2, \ldots, n-1$, then $A_\alpha\,dx^\alpha$ is integrable. Thus, $\exists$ functions $\overline{M}(x^\alpha; x^n)$ and $\overline{\phi}(x^\alpha; x^n)$ such that $\overline{M}A_\alpha\,dx^\alpha = d\overline{\phi}$ where x^n is regarded as a parameter. But $\overline{\phi}$ regarded as a function of n variables satisfies $d\overline{\phi} = (\partial\overline{\phi}/\partial x^\alpha)\,dx^\alpha + (\partial\overline{\phi}/\partial x^n)\,dx^n = \overline{M}A_\alpha\,dx^\alpha + (\partial\overline{\phi}/\partial x^n)\,dx^n$, and therefore $\overline{M}\,\vec{d}q - d\overline{\phi} = (\overline{M}A_n - \partial\overline{\phi}/\partial x^n)\,dx^n \equiv \Delta\,dx^n$. It is easily proven that if the A_k satisfy the integrability conditions, then so too will the functions obtained from A_k by multiplication by any differentiable function. Hence, if we write $D_k = \overline{M}A_k$, then $D_i(\partial D_k/\partial x^j - \partial D_j/\partial x^k) + D_j(\partial D_i/\partial x^k - \partial D_k/\partial x^i) + D_k(\partial D_j/\partial x^i - \partial D_i/\partial x^j) = 0$. If we now set $i = n$, $j = \alpha$, $k = \beta$ and use $\partial\overline{\phi}/\partial x^\alpha = \overline{M}A_\alpha$, this becomes

$\overline{M}A_n(\partial^2\overline{\phi}/\partial x^\alpha\,\partial x^\beta - \partial^2\overline{\phi}/\partial x^\beta\,\partial x^\alpha) + \partial\overline{\phi}/\partial x^\alpha[\partial(\overline{M}A_n)/\partial x^\beta$
$- \partial^2\overline{\phi}/\partial x^n\,\partial x^\beta] + \partial\overline{\phi}/\partial x^\beta[\partial^2\overline{\phi}/\partial x^n\,\partial x^\alpha - \partial(MA_n)/\partial x^\alpha] = 0.$
The first term is identically zero, while the remainder
can be rewritten as $(\partial\overline{\phi}/\partial x^\alpha)(\partial\Delta/\partial x^\beta) - (\partial\overline{\phi}/\partial x^\beta)(\partial\Delta/\partial x^\alpha)$
$= \partial(\overline{\phi}, \Delta)/\partial(x^\alpha, x^\beta) = 0$ ∀ $\alpha, \beta = 1, 2, \ldots, n-1$. But, by
Corollary III.3, $\overline{\phi}$ and Δ are dependent insofar as their
dependence on x^α, and so at worst $\Delta = \Delta(\overline{\phi}, x^n)$. Hence,
$dg \equiv d\overline{\phi} + \Delta dx^n$ is a differential form in two variables,
$\overline{\phi}$ and x^n. Let $\Phi(\overline{\phi}, x^n) = \text{const}$ be a solution of $dg = 0$,
that is, of the differential equation $d\overline{\phi}/dx^n = -\Delta(\overline{\phi}, x^n)$.
Then $dg = 0$ iff $d\Phi = 0$. But as in Corollary III.5 this
condition $\Rightarrow$ the existence of a function $\lambda = \lambda(\overline{\phi}, x^n)$ such
that $\lambda dg = d\Phi$ and so dg is integrable. But then $d\Phi$
$= \lambda dg = \lambda(\overline{M} dq)$ and thus $\lambda\overline{M}$ is an integrating factor for
dq; hence, by induction dq is integrable for any n. That
an integrating factor cannot be unique is obvious, since
if M is an integrating factor, so too is kM, where k is a
constant. As a matter of fact, any function of the
pseudopotential, corresponding to M, when multiplied
by M will give an integrating factor.

There is no *a priori* reason why the region of in-
tegrability V of dq should in any way be connected with
the chain of $(\Sigma, \leq_A, =_A)$. Is it possible, working solely in
V, to determine if V can be thought of as the image of
some chain? This certainly is not possible without im-
posing some conditions because the relation $\leq_A$ is
defined in terms of adiabatic processes which need not
be quasistatic, whereas if we work only in V, then we
can carry out only quasistatic processes. However,
Corollary III.6 offers a clue as to just what the addi-
tional conditions must be.

*Theorem III.*10: Suppose $V \subset \text{Im} e(\mathcal{T})$ is an open, con-
nected subset, $C = [e(\mathcal{T})]^{-1}(V) = \{x \in \Sigma \,|\, e(\mathcal{T})(x) \in V\}$, dq
integrable on V, M an integrating factor, and ϕ the cor-
responding pseudopotential. If $x \leq_A \overline{x}$ iff ∃ some process
P such that $x \xrightarrow{P} \overline{x}$, $\text{Im} P \subset C$, and $q(P)\,\text{sgn}(M) \geq 0$, then C
is a chain in $(\Sigma, \leq_A, =_A)$ and ϕ is the coordinate rep-
resentation of an entropy for C.

Proof: Let φ be the function on C whose coordinate
representation is ϕ, that is, $\varphi = \phi \circ e(\mathcal{T}) | C$. Then just
as in the proof of Corollary III.5 it can be readily
established that

$$[\varphi(\overline{x}) - \varphi(x)]/|M|_{\text{max}}$$
$$\leq q(P)\,\text{sgn}(M) \leq [\varphi(\overline{x}) - \varphi(x)]/|M|_{\text{min}}$$

∀ processes P such that $x \xrightarrow{P} \overline{x}$ and $\text{Im} P \subset C$. From this it
follows that $\varphi(x) \leq \varphi(\overline{x})$ iff $q(P)\,\text{sgn}(M) \geq 0$ ∀ processes
P such that $x \xrightarrow{P} \overline{x}$ and $\text{Im} P \subset C$, and also if $q(P)\,\text{sgn}(M)$
≥ 0 for one such process, then $q(P)\,\text{sgn}(M) \geq 0$ for all
such processes. Note that no conclusion can be drawn
concerning the quasistatic heat effect for quasistatic
processes for which $\text{Im} P \not\subset C$. Now since V is an open,
connected subset of R^n, then V is pathwise connected
(W, p. 199, Cor. 27.6) and hence C is pathwise con-
nected since C and V are homeomorphic images of each
other under $e(\mathcal{T}) | C$. Thus, for each pair $x, \overline{x} \in C$ ∃ some
process P linking x and $\overline{x}, x \xrightarrow{P} \overline{x}$ and $\text{Im} P \subset C$. Now either
$q(P)\,\text{sgn}(M) \geq 0 \Rightarrow x \leq_A \overline{x}$ and $\varphi(x) \leq \varphi(\overline{x})$, or $q(P)\,\text{sgn}(M)$
$< 0 \Rightarrow q(P^*)\,\text{sgn}(M) > 0 \Rightarrow q(P^*)\,\text{sgn}(M) \geq 0 \Rightarrow \overline{x} \leq_A x$ and
$\varphi(\overline{x}) \leq \varphi(x)$. Thus, ∀ $x, \overline{x} \in C$, $x \leq_A \overline{x}$ or $\overline{x} \leq_A x$, and

hence C is a chain. Further, $x =_A \overline{x}$ iff $\varphi(x) \leq \varphi(\overline{x})$, and
since $x =_A \overline{x}$ iff $x \leq_A \overline{x}$ and $\overline{x} \leq_A x$ iff $\varphi(x) \leq \varphi(\overline{x})$ and
$\varphi(\overline{x}) \leq \varphi(x)$ iff $\varphi(x) = \varphi(\overline{x})$, and hence φ is an entropy
and ϕ is its coordinate representation.

A connection between pseudopotential and empirical
entropy cannot be made unless one can verify the con-
dition on $\leq_A$ contained in Theorem III.10 or else one is
prepared to accept it as an assumption. In either event
its function is analogous to that of Axiom I.54 where
the relation $=_A$ was expressed in terms of heat. Here
the relation $\leq_A$, and hence also $=_A$, is related to quasi-
static heat. The surprising feature of this latter con-
nection is that it involves $\text{sgn}(M)$. But a little reflection
dispels this surprise for $\text{sgn}(M)$ is arbitrary because,
if M is an integrating factor, so too is $(-M)$. Con-
sequently, the condition $q(P)\,\text{sgn}(M) \geq 0$ really is nothing
more than the requirement that $\text{sgn}[q(P)]$ be invariant for
all processes P such that $\text{Im} P \subset C$. The factor $\text{sgn}(M)$
was retained in Theorem III.10 merely to reinforce the
warning that Theorem III.10 critically depends upon
dq being integrable. Without integrability there would
be no theorem.

*Corollary III.*11: Suppose Theorem III.10 is valid and
ϕ is the pseudopotential corresponding to the integrat-
ing factor M for dq. If V is of dimension $\overline{n}$ with intrinsic
coordinates $z^1, z^2, \ldots, z^{\overline{n}}$, then in V we have $\partial\phi/\partial z^\alpha$
$= MA_i\,\partial x_i/\partial z^\alpha$, $\alpha = 1, 2, \ldots, \overline{n}$. If $\overline{n} = n$, then $\partial\phi/\partial u = M$,
$\partial\phi/\partial x^i = MA_i$, $i = 1, 2, \ldots, n$, and $A_i = -(\partial u/\partial x^i)_\varphi$,
$i = 2, 3, \ldots, n$.

Proof: See corresponding proof in Corollary III.5.

Several features of thermodynamics remains to be
clarified. These include (1) a determination of the num-
ber of independent integrating factors and pseudopoten-
tials, (2) a consideration of the role played by the in-
tensive—extensive partitioning of thermodynamic
variables, and (3) an exploration of the connection be-
tween integrating factors and empirical temperatures.
I shall now look at the first of these three items. This
question is examined by Forsyth (p. 3).[21] However, his
discussion is predicated on the assumption that only one
relation connects the differentials dx^i. Clearly, this is
an assumption which goes beyond any statement con-
fined to properties of dq alone. The next theorem de-
pends only upon dq and the fact that it is integrable.

*Theorem III.*12: Let V be an open, connected subset
of $\text{Im} e(\mathcal{T})$, $dq = A_k dx^k$ is integrable on V and also the
functions $A_k = A_k(x^i) \neq 0$, $k = 1, 2, \ldots, n$, are of class
$C^{(1)}$ on V and, for some k, j, $\partial A_k/\partial x^j - \partial A_j/\partial x^k \neq 0$. Then
(1) any two pseudopotentials are dependent, and (2) if M
is an integrating factor and ϕ the corresponding
pseudopotential, then M and ϕ are independent. Further,
(3) dq possesses only two independent integrating fac-
tors, and (4) if M is an integrating factor and ϕ the cor-
responding pseudopotential, then any other integrating
factor is of the form $\overline{M} = Mg(\phi)$, where $g(\phi)$ is any
function of ϕ and $\overline{M}$ is independent of M iff g is not a
constant function.

Proof: To establish (1) and (2) suppose ϕ_1, ϕ_2 are
pseudopotential corresponding to the integrating factors
M_1, M_2. Then $\partial\phi_1/\partial x^k = M_1 A_k$, $\partial\phi_2/\partial x^j = M_2 A_j$, and hence
$\partial(\phi_1, \phi_2)/\partial(x^k, x^j) = (\partial\phi_1/\partial x^k)(\partial\phi_2/\partial x^j) - (\partial\phi_1/\partial x^j)(\partial\phi_2/\partial x^k)$

$= (M_1 A_k)(M_2 A_j) - (M_1 A_j)(M_2 A_k) \equiv 0$. Now consider
$\partial(\phi, M)/\partial(x^k, x^j) = (\partial\phi/\partial x^k)(\partial M/\partial x^j) - (\partial\phi/\partial x^j)(\partial M/\partial x^k)$
$= M(A_k \, \partial M/\partial x^j - A_j \, \partial M/\partial x^k) = M^2(\partial A_j/\partial x^k - \partial A_k/\partial x^j)$,
where the last equality comes from the differential
equation, satisfied by M, given in Theorem III. 9.
Obviously, since $M \neq 0$, $\partial(\phi, M)/\partial(x^k, x^j) = 0$ iff $\partial A_j/\partial x^k$
$- \partial A_k/\partial x^j = 0$. But by hypothesis this is nonzero for
some k, j, and, by Corollary III. 3, ϕ and M are inde-
pendent, and ϕ_1 and ϕ_2 are dependent. The property (3)
is established by supposing M_λ, $\lambda = 1, 2, 3$, to be in-
tegrating factors. Since dq is integrable, at least one,
say M_2, is nontrivial, that is, $M_2 \neq 0$. By Theorem
III. 9 we know that $A_k \, \partial M_\lambda/\partial x^j - A_j \, \partial M_\lambda/\partial x^k = M_\lambda(\partial A_j/\partial x^k$
$- \partial A_k/\partial x^j)$, $\lambda = 1, 2, 3$. Now consider the expression
$A_k \, \partial(M_\lambda/M_2)/\partial x^j - A_j \, \partial(M_\lambda/M_2)/\partial x^k = (1/M_2)(A_k \, \partial M_\lambda/\partial x^j$
$- A_j \, \partial M_\lambda/\partial x^k) - (M_\lambda/M_2^2)(A_k \, \partial M_2/\partial x^j - A_j \, \partial M_2/\partial x^k)$
$= (M_\lambda/M_2)(\partial A_j/\partial x^k - \partial A_k/\partial x^j) - (M_\lambda M_2/M_2^2)(\partial A_j/\partial x^k$
$- \partial A_k/\partial x^j) \equiv 0$. Hence, with $\lambda = 1, 3$ we have
$A_k \, \partial(M_1/M_2)/\partial x^j - A_j \, \partial(M_1/M_2)/\partial x^k = 0$ and
$A_k \, \partial(M_3/M_2)/\partial x^j - A_j \, \partial(M_3/M_2)/\partial x^k = 0$. Since $A_i \neq 0$,
$i = 1, 2, \ldots, n$, the A_i are a nontrivial solution to a pair
of homogeneous, linear equations. But homogeneous,
linear equations have a nontrial solution iff the deter-
minant vanishes. Hence, $\partial(M_1/M_2, M_3/M_2)/\partial(x^k, x^j) = 0$
$\forall \, k, j$. But by Corollary III. 3 and Definition III. 1(2) we
have $\Psi(M_1/M_2, M_3/M_2) = 0$ and M_1, M_2, M_3 are dependent.
Thus, there can be no more than two independent in-
tegrating factors. Suppose $\bar{M} = g(\phi)M$; then by direct
computation $\partial(\bar{M}, M)/\partial(x^k, x^j) = (dg/d\phi)\partial(\phi, M)/\partial(x^k, x^j)$.
If $dg/d\phi \neq 0$, then $\bar{M}, M$ are independent since ϕ and M
are. Hence, there always exist two independent in-
tegrating factors since it is trivial to verify that $\bar{M}$ is
an integrating factor. This completes the proof of (3)
and a portion of (4). To complete the proof of (4), we
consider a third integrating factor $\tilde{M}$. By (3) $\exists$ a func-
tion Ψ such that $\tilde{M} = \Psi(M, \bar{M})$. Now by differentiation
$\partial\tilde{M}/\partial x^k = (\partial\Psi/\partial M)(\partial M/\partial x^k) + (\partial\Psi/\partial\bar{M})(g(\phi) \, \partial M/\partial x^k$
$+ \dot{g}M \, \partial\phi/\partial x^k)$, where $\dot{g} = dg/d\phi$. Substituting into the
differential equation satisfied by $\tilde{M}$ we have, using
$\partial\phi/\partial x^k = MA_k$, $0 = A_k \, \partial\tilde{M}/\partial x^j - A_j \, \partial\tilde{M}/\partial x^k - \tilde{M}(\partial A_j/\partial x^k$
$- \partial A_k/\partial x^j) = (\partial\Psi/\partial M + g \, \partial\Psi/\partial\bar{M})(A_k \, \partial M/\partial x^j - A_j \, \partial M/\partial x^k)$
$- \Psi(\partial A_j/\partial x^k - \partial A_k/\partial x^j) = [(\partial\Psi/\partial M + g \, \partial\Psi/\partial\bar{M})M - \Psi]$
$\times (\partial A_j/\partial x^k - \partial A_k/\partial x^j)$. Since $\partial A_j/\partial x^k - \partial A_k/\partial x^j \neq 0$ and
$gM = \bar{M}$, we find that Ψ is a solution of $M \, \partial\Psi/\partial M$
$+ \bar{M} \, \partial\Psi/\partial\bar{M} = \Psi$. The general integral of this equation is
obtained from two independent solutions of the sub-
sidiary equations $dM/M = d\bar{M}/\bar{M} = d\Psi/\Psi$. The first equal-
ity has the solution $\bar{M}/M = c_1$ where c_1 is a constant,
while by equating the first and third terms we get the
solution $\Psi/M = c_2$ where c_2 is a constant. The general
integral is thus of the form $\Psi/M = G(\bar{M}/M)$. But $\bar{M}/M$
$= g(\phi)$; hence, $\Psi = MG(g(\phi))$ and the theorem has been
proven.

Extensive and intensive variables are endemic to
thermodynamics, and this partitioning of physically
relevant quantities does not seem to extend to other
physical theories. Roughly speaking, extensive varia-
bles are often said to be proportional to the mass of the
system while intensive variables are said to be inde-
pendent of the mass. Alternatively, it is sometimes
said that extensive variables are additive over the sub-
systems of a composite system. The first of these has
no meaning in a local theory, while the second has no

place in a theory which avoids composite systems. If
we are to explore the consequences of such a partition-
ing of thermodynamic variables, then we must find an
alternative, but equivalent, expression of the exten-
sive—intensive property. The extensive—intensive
partitioning reflects physically significant properties of
these variables. Further, mathematics and physics
differ in their treatment of functions only because
physics insists that each function bear a set of physical
units or dimensions. Hence, the obvious place to look
for the extensive—intensive characterization is in the
physical units assigned to functions. The following
definition does just this for real-valued functions on Σ,
although it could equally well apply to real-valued func-
tions on X.

Definition III. 13. Suppose $\Psi : \Sigma \to R$ is a real-valued
function. The function Ψ is said to be extensive iff Ψ is
a specific quantity, that is, Ψ is expressed per unit
mass. If Ψ is not extensive, then it is said to be inten-
sive. Let $\overline{T}$ be the collection of Axiom II. 1, and suppose
$\bar{n}$ members of $\overline{T}$, say $\tau^1, \tau^2, \ldots, \tau^{\bar{n}}$, are extensive and
the $n - \bar{n}$ members $\tau^{\bar{n}+1}, \ldots, \tau^n$ are intensive, where $\bar{n}$
is an integer and $0 \leq \bar{n} \leq n$. Then $\overline{T}$ is said to be exten-
sive of rank $\bar{n}$. If $\overline{T}$ is extensive of rank $\bar{n}$, then the
following index convertion will be adopted: (1) Latin
indices will use the range $1, 2, \ldots, n$. (2) Greek indices
from the first part of the alphabet $(\alpha, \beta, \gamma, \cdots)$ will use
the range $1, 2, \ldots, \bar{n}$. (3) Greek indices from the last
part of the alphabet $(\mu, \nu, \sigma, \tau, \cdots)$ will use the range
$\bar{n} + 1, \ldots, n$.

Axiom III. 14: Suppose $\Psi : \Sigma \to R$ is a real-valued func-
tion and $\psi(x^k) = \psi(x^\alpha, x^\mu)$ is its coordinate representation
in Σ. Then (1) Ψ is extensive iff $\psi(x^k)$ is homogeneous
of degree one in the extensive variables, that is,
$\psi(\lambda x^\alpha, x^\mu) = \lambda\psi(x^\alpha, x^\mu)$. (2) Ψ is intensive iff $\psi(x^k)$ is
homogeneous of degree zero in the extensive variables,
that is, $\psi(\lambda x^\alpha, x^\mu) = \psi(x^\alpha, x^\mu)$.

Definition III. 13 might be regarded as the physical
characterization of extensive and intensive quantities,
while Axiom III. 14 serves to relate this physical char-
acterization of a function to the mathematical charac-
terization of homogeneity. It would be nice to replace
Axiom III. 14 by a corresponding theorem, but there
does not seem to be an obvious proof. In the absence of
such a proof Axiom III. 14 serves to limit the class of
thermodynamic functions. It represents a reasonable
restriction since, at least for algebraic functions, it is
sufficient to guarantee that the physical units of ψ will
be correct.

Some clarifying remarks are necessary before we can
begin an analysis of the significance of the intensive-
extensive partitioning of variables. The form $A_k(x^i) \, dx^k$
has been indicated simply by dq, and no attempt was
made to indicate that its evaluation required the x^i co-
ordinates and their differentials. This created no prob-
lem since essentially only the x^i coordinates were ever
used. We will shortly consider the form $A_k(z^i) \, dz^k$,
where the z^i are new coordinates related to the x^i by a
particular kind of coordinate transformation $z^i = z^i(x^k)$.
Whenever it becomes necessary to distinguish between
the two differential forms, we will write $dq_x = A_k(x^i) \, dx^k$
and $dq_z = A_k(z^i) \, dz^k$. In general, we cannot expect dq_x

to equal dq_z despite the fact that the same functions were used in both cases.

Definition III. 15: The differential form dq is said to be extensive on $\mathrm{Im}e(\mathcal{T})$ iff $dq_z = \lambda\, dq_x$ for the coordinate transformation $z^\alpha = \lambda x^\alpha$, $z^\mu = x^\mu$, where $\lambda \neq 0$ is independent of x^i and $dq_z = A_k(z^i)\,dz^k$, $dq_x = A_k(x^i)\,dx^k$.

Theorem III. 16: The differential form dq is extensive on $\mathrm{Im}e(\mathcal{T})$ iff the functions A_α are intensive and the functions A_μ are extensive. That is, iff $A_\alpha(\lambda x^\beta, x^\nu) = A_\alpha(x^\beta, x^\nu)$ and $A_\mu(\lambda x^\beta, x^\nu) = \lambda A_\mu(x^\beta, x^\nu)$.

Proof: Since $dq_z = A_\alpha(z^\beta, z^\nu)\,dz^\alpha + A_\mu(z^\beta, z^\nu)\,dz^\mu$
$= \lambda A_\alpha(\lambda x^\beta, x^\nu)\,dx^\alpha + A_\mu(\lambda x^\beta, x^\nu)\,dx^\mu$, it follows that dq_z
$-\lambda\, dq_x = \lambda[A_\alpha(\lambda x^\beta, x^\nu) - A_\alpha(x^\beta, x^\nu)]\,dx^\alpha + [A_\mu(\lambda x^\beta, x^\nu)$
$-\lambda A_\mu(x^\beta, x^\nu)]\,dx^\mu$. But in $\mathrm{Im}e(\mathcal{T})$ the x^k are independent and hence, since $\lambda \neq 0$, the theorem follows.

Relatively important, for the structure of the thermodynamics based on the use of a pseudopotential as an entropy is the assertion that the pseudopotential is extensive. Yet, so far as I know, no one has questioned the validity of this assumption or attempted to prove it. I shall now deal with just this question. Is it always possible to choose the pseudopotential as extensive and, concomitantly, the integrating factor as intensive?

Theorem III. 17: Let $dq = A_k(x^i)\,dx^k$ be extensive on $\mathrm{Im}e(\mathcal{T})$, integrable on an open, connected subset $V \subset \mathrm{Im}e(\mathcal{T})$ with A_k of class $C^{(1)}$ on V. If $M(x^\alpha, x^\mu)$ is an integrating factor on V, then for any constant $\lambda \neq 0$, $M(\lambda x^\alpha, x^\mu)$ is an integrating factor.

Proof: Since dq is integrable, $\exists$ an integrating factor M, which by Theorem III. 9 satisfies $A_i(x^k)\,\partial M/\partial x^j$
$-A_j(x^k)\,\partial M/\partial x^i = M(\partial A_j/\partial x^i - \partial A_i/\partial x^j)$. There are three cases which must be considered. For $i = \alpha$, $j = \beta$ we have $A_\alpha(x^\gamma, x^\nu)\,\partial M/\partial x^\beta - A_\beta(x^\gamma, x^\nu)\,\partial M/\partial x^\alpha$
$= M[\partial A_\beta(x^\gamma, x^\nu)/\partial x^\alpha - \partial A_\alpha(x^\gamma, x^\nu)/\partial x^\beta]$. Dividing by λ and using Theorem III. 16 we obtain $A_\alpha(z^\gamma, z^\nu)\,\partial M/\partial z^\beta$
$-A_\beta(z^\gamma, z^\nu)\,\partial M/\partial z^\alpha = M[\partial A_\beta(z^\gamma, z^\nu)/\partial z^\alpha - \partial A_\alpha(z^\gamma, z^\nu)/\partial z^\beta]$.
In a similar manner we find for $i = \alpha$, $j = \mu$ and $i = \sigma$, $j = \tau$ the results $A_\alpha(z^\gamma, z^\nu)\,\partial M/\partial z^\mu - A_\mu(z^\gamma, z^\nu)\,\partial M/\partial z^\alpha$
$= M[\partial A_\mu(z^\gamma, z^\nu)/\partial z^\alpha - \partial A_\alpha(z^\gamma, z^\nu)/\partial z^\mu]$ and $A_\sigma(z^\gamma, z^\nu)$
$\times \partial M/\partial z^\tau - A_\tau(z^\gamma, z^\nu)\,\partial M/\partial z^\sigma = M[\partial A_\tau(z^\gamma, z^\nu)/\partial z^\sigma$
$-\partial A_\sigma(z^\gamma, z^\nu)/\partial z^\tau]$. Hence, $A_i(z^k)\,\partial M/\partial z^j - A_j(z^k)\,\partial M/\partial z^i$
$= M(\partial A_j/\partial z^i - \partial A_i/\partial z^j)$. Thus, if $M(x^\alpha, x^\mu)$ is an integrating factor, so too is $M(z^\alpha, z^\mu) = M(\lambda x^\alpha, x^\mu)$.

This theorem says, in effect, that the transformation $z^\beta = \lambda x^\beta$, $z^\mu = x^\mu$ is a symmetry transformation for the system of partial differential equations whose solutions are the integrating factors for dq. It thus guarantees the existence of a continuous one parameter family of integrating factors $M(\lambda x^\beta, x^\mu)$. But by Theorem III. 12(3) only two members of this family can be independent.

Theorem III. 18: Let $dq = A_k(x^i)\,dx^k$ be extensive on $\mathrm{Im}e(\mathcal{T})$, integrable on an open connected subset $V \subset \mathrm{Im}e(\mathcal{T})$ with $A_k \neq 0$ of class $C^{(1)}$ on V and $\partial A_k/\partial x^j$
$-\partial A_j/\partial x^k \neq 0$ for some k, j. If $M(x^\alpha, x^\mu)$ is an integrating factor on V, then M is homogeneous of degree p in the extensive variables for some real p. That is, $M(\lambda x^\alpha, x^\mu) = \lambda^p M(x^\alpha, x^\mu)$ and p real.

Proof: Write $M_\lambda(x^\alpha, x^\mu) = M(\lambda x^\alpha, x^\mu)$. Then $M_1(x^\alpha, x^\mu)$
$= M(x^\alpha, x^\mu)$. Let ϕ be the pseudopotential corresponding

to M. Then since M_λ is an integrating factor, by Theorem III. 17 it follows from Theorem III. 12(4) that M_λ
$= Mg(\phi; \lambda)$. Differentiation with respect to ϕ gives $\partial M_\lambda/\partial\phi = M\,\partial g/\partial\phi + g\,\partial M/\partial\phi$. But $\partial M_\lambda/\partial\phi = 0 = \partial M/\partial\phi$, and since $M \neq 0$, we conclude that $\partial g/\partial\phi = 0$ and hence $M_\lambda = Mg(\lambda)$. If this result is differentiated with respect to λ to get $[\partial M_\lambda/\partial(\lambda x^\beta)]\,\partial(\lambda x^\beta)/\partial\lambda = M\dot{g}(\lambda)$ where $\dot{g}(\lambda)$
$= dg/d\lambda$, λ is set equal to 1, and in place of $\dot{g}(1)$ we write p, we then find that $x^\beta\,\partial M/\partial x^\beta = pM$. Hence, for some real number p, M is a solution of a partial differential equation whose subsidiary equations are dx^1/x^1
$= dx^2/x^2 = \cdots = dx^{\bar{n}}/x^{\bar{n}} = dM/pM$. Now $\bar{n}$ independent integrals of these equations are easily found by selecting one of the x^α, say x^1, and solving $dx^1/x^1 = dx^\alpha/x^\alpha$, $\alpha = 2, 3, \ldots, \bar{n}$, and $dx^1/x^1 = dM/pM$. The solutions are $x^\alpha/x^1 = c^\alpha$, $\alpha = 2, 3, \ldots, \bar{n}$, and $M/(x^1)^p = c$, where c and c^α are constants. The general integral of the partial differential equation for M is $M = (x^1)^p\Psi(x^2/x^1, \ldots, x^{\bar{n}}/x^1, x^\mu)$. This form immediately shows that M is homogeneous of degree p in the extensive variables where p is real.

Corollary III. 19: Let $dq = A_k(x^i)\,dx^k$ be extensive on $\mathrm{Im}e(\mathcal{T})$, integrable on an open, connected subset $V \subset \mathrm{Im}e(\mathcal{T})$ with $A_k \neq 0$ of class $C^{(1)}$ on V and $\partial A_k/\partial x^j$
$-\partial A_j/\partial x^k \neq 0$ for some k, j, $M(x^\alpha, x^\mu)$ an integrating factor for dq on V, homogeneous of degree p in the extensive variables, p real and ϕ the corresponding pseudopotential. (1) If $(p+1) \neq 0$, then ϕ can be chosen to be homogeneous of degree $(p+1)$ in the extensive variables. (2) If $(p+1) = 0$, then for some $x^\gamma \neq 0$, ϕ has the form $\phi = c\,\ln x^\gamma + \Psi$, where Ψ is a function homogeneous of degree zero in the extensive variables with the form $\Psi(x^1/x^\gamma, \ldots, x^{\gamma-1}/x^\gamma, x^{\gamma+1}/x^\gamma, \ldots, x^{\bar{n}}/x^\gamma, x^\mu)$, and c is a constant. The constant $c = 0$ iff $x^\alpha A_\alpha = 0$, and iff $c \neq 0$, then $c = Mx^\alpha A_\alpha$ and M can always be chosen so that $c = 1$.

Proof: Suppose $M(z^\alpha, z^\mu) = M(\lambda x^\alpha, x^\mu) = \lambda^p M(x^\alpha, x^\mu)$. Then $d\phi(z^i) = M(z^i)\,dq_z = [\lambda^p M(x^i)][\lambda\, dq_x] = \lambda^{p+1}\,d\phi(x^i)$. Hence, $d[\phi(z^i) - \lambda^{p+1}\phi(x^i)] = 0$ or $\phi(z^i) = \lambda^{p+1}\phi(x^i) + c(\lambda)$, where $c(\lambda)$ is at most a function of λ. Thus, we have $\phi(\lambda x^\alpha, x^\mu) = \lambda^{p+1}\phi(x^\alpha, x^\mu) + c(\lambda)$ and differentiating with respect to λ gives $x^\beta\,\partial\phi(\lambda x^\alpha, x^\mu)/\partial(\lambda x^\beta) = (p+1)\lambda^p$
$\times \phi(x^\alpha, x^\mu) + \dot{c}(\lambda)$, where $\dot{c} = dc/d\lambda$. Letting $\lambda = 1$, we find that ϕ satisfies the partial differential equation $x^\beta\,\partial\phi/\partial x^\beta = (p+1)\phi + \dot{c}(1)$ with subsidiary equations dx^1/x^1
$= \cdots = dx^{\bar{n}}/x^{\bar{n}} = d\phi/[(p+1)\phi + \dot{c}(1)]$. For some fixed index γ this system has $(n-1)$ independent integrals $x^\alpha/x^\gamma = c^\alpha$ with $\alpha \neq \gamma$. The remaining independent integral takes one of two forms depending upon whether or not $(p+1) = 0$. If $(p+1) \neq 0$, then $dx^\gamma/x^\gamma = d\phi/[(p+1)\phi + \dot{c}(1)]$ has a solution $[(p+1)\phi + \dot{c}(1)]/(x^\gamma)^{p+1} = c_1$, while if $(p+1) = 0$, we have the solution $\phi - \dot{c}(1)\ln x^\gamma = c_2$. The general integral has the form $\phi = (x^\gamma)^{p+1}\Psi(x^\alpha/x^\gamma, \alpha \neq \gamma, x^\mu)/(p+1) - \dot{c}(1)/(p+1)$ for $(p+1) \neq 0$ and $\phi = \dot{c}(1)\ln x^\gamma + \Psi(x^\alpha/x^\gamma, \alpha \neq \gamma, x^\mu)$ for $(p+1) = 0$. If ϕ is a pseudopotential, $\phi + k$ is a psuedopotential for any constant k. In the case $(p+1) \neq 0$, the choice $k = \dot{c}(1)/(p+1)$ gives a psuedopotential which is homogeneous of degree $(p+1)$ in the extensive variables for Ψ is obviously homogeneous of degree zero in the extensive variables. In the case $(p+1) = 0$ we have by differentiations $MA_\alpha = \partial\phi/\partial x^\alpha = \delta^\gamma_\alpha\dot{c}(1)/x^\gamma + \partial\Psi/\partial x^\alpha$. But since Ψ is homogeneous of degree zero in the extensive variables

we have $x^\alpha \partial \Psi / \partial x^\alpha = 0$ by Euler's theorem on homogeneous functions. And hence since $x^\alpha \delta_\alpha^\gamma = x^\gamma$, we find $M A_\alpha x^\alpha = \overset{\circ}{c}(1)$. Now $M \neq 0$, and therefore $\overset{\circ}{c}(1) = 0$ iff $x^\alpha A_\alpha = 0$. To complete the proof observe that if M is an integrating factor, then $\bar{M} = kM$ is an integrating factor for any constant k. The choice $k^{-1} = \overset{\circ}{c}(1)$ gives $\bar{M} A_\alpha x^\alpha = 1$ or $\bar{M}^{-1} = A_\alpha x^\alpha$.

Theorem III.20: Let $dq = A_k(x^i) dx^k$ be extensive on $\mathrm{Im}e(T)$, integrable on an open, connected subset $V \subset \mathrm{Im}e(T)$ with $A_k \neq 0$, $k = 1, 2, \ldots, n$, of class $C^{(1)}$ on V, $\partial A_k / \partial x^j - \partial A_j / \partial x^k \neq 0$ for some k, j, $x^\alpha A_\alpha \neq 0$, $M(x^\alpha, x^\mu)$ an integrating factor and ϕ the corresponding pseudopotential. Then $x^\alpha A_\alpha$ is not a constant function, M can always be chosen to be homogeneous of degree $p = 0$ or $p = -1$ in the extensive variables, and $p = -1$ iff $(x^\alpha A_\alpha)^{-1}$ is an integrating factor. If $p = 0$, then $\phi / M = x^\alpha A_\alpha$.

Proof: Suppose M is homogeneous of degree $p \neq -1$. Then by Corollary III.19(1) the pseudopotential ϕ can be chosen to be homogeneous of degree $(p+1)$ in the extensive variables, and since $(-M)$ is an integrating factor with pseudopotential $(-\phi)$, it is no restriction to assume $\mathrm{sgn}(\phi) = 1$. By Theorem III.12(4) $\bar{M} = M / [(p+1)\phi^{p/(p+1)}]$ is an integrating factor. Further, $\bar{M}(z^i) = M(z^i) / [(p+1)\phi^{p/(p+1)}(z^i)] = [\lambda^p / (\lambda^{p+1})^{p/(p+1)}] M(x^i) / [(p+1)\phi^{p/(p+1)}(x^i)] = \lambda^0 \bar{M}(x^i)$. Thus, $\bar{M}$ is intensive and the corresponding pseudopotential $\bar{\phi}$ is extensive by Corollary III.19(1). In fact, $d\bar{\phi} = \bar{M} dq = d\phi / [(p+1)\phi^{p/(p+1)}(x)] = d\phi^{1/(p+1)}$. Thus, it is no restriction to assume $p = 0$ or $p = -1$. By Corollary III.19(2), $p = -1$ implies that $(x^\alpha A_\alpha)^{-1}$ is an integrating factor, and thus if $(x^\alpha A_\alpha)^{-1}$ is not an integrating factor, $p \neq -1$. Hence, $p = -1$ iff $(x^\alpha A_\alpha)^{-1}$ is an integrating factor, and a constant integrating factor contradicts $\partial A_k / \partial x^j - \partial A_j / \partial x^k \neq 0$. If $p = 0$, we know by Corollary III.19(1) that ϕ is extensive, and by Euler's theorem on homogenous functions $\phi = x^\alpha \partial \phi / \partial x^\alpha = x^\alpha M A_\alpha$ or $\phi / M = x^\alpha A_\alpha$. If $x^\alpha A_\alpha$ is a constant function, then ϕ and M are dependent, and this contradicts Theorem III.12(2).

Validation of conventional thermodynamics can now be seen as the justification of the conditions which are required for the applicability of Theorem III.9 (dq is integrable), Theorem III.10 (the pseudopotential is an entropy), and Theorem III.20 (the pseudopotential is an extensive function). But to accomplish this validation it is necessary that one possess the expressions for the functions $A_k(x^i)$ so that the criteria set forth for the applicability of these theorems may be tested. There is one additional aspect of the usual thermodynamics which needs examination. This is the connection between the coordinate representation $\theta(x^i)$ of an empirical temperature Θ and an integrating factor M for dq. In the language of Theorem II.12 this could be phrased as establishing a connection between the absolute temperature T_ϕ, associated with an entropy ϕ, and θ. Obviously, if $T_\phi d\phi = dq$ and $d\phi = M dq$ then $T_\phi = M^{-1}$. If such a connection can be established, then a nonmeasurement T_ϕ is endowed with experimental significance by virtue of its association with a measurement Θ. The usual thermodynamics asserts that a linear relationship exists between T_ϕ and some θ. This

linear relationship is merely a special case of functional dependence and, since there is no particular significance to linearity, I shall deal with the less restrictive case.

Theorem III.21: Let $\mathcal{J}$ be a collection of real valued, nonconstant functions on Σ whose coordinate representations are of class $C^{(1)}$. Suppose $\Psi_1, \Psi_2 \in \mathcal{J}$ and $\psi_1(x^i)$, $\psi_2(x^i)$ are the respective coordinate representations. A relation $\overset{T}{\approx}$ is defined on $\mathcal{J}$ by $\Psi_1 \overset{T}{\approx} \Psi_2$ iff $\psi_1(x^i)$ and $\psi_2(x^i)$ are functionally dependent. The relation $\overset{T}{\approx}$ is an equivalence relation on $\mathcal{J}$. If $\Psi_1 \overset{T}{\approx} \Psi_2$, then Ψ_1 and Ψ_2 are said to be thermodynamically equivalent.

Proof: The relation $\overset{T}{\approx}$ is reflexive since $\psi_1 - \psi_1 = 0$ shows $\Psi_1 \overset{T}{\approx} \Psi_1$ and symmetric since $\Psi_1 \overset{T}{\approx} \Psi_2 \Rightarrow g(\psi_1, \psi_2) = 0 \Rightarrow \Psi_2 \overset{T}{\approx} \Psi_1$. To show transistivity consider $\Psi_1 \overset{T}{\approx} \Psi_2$ and $\Psi_2 \overset{T}{\approx} \Psi_3$. By Corollary III.3 this $\Rightarrow \partial(\psi_1, \psi_2) / \partial(x^i, x^j) = 0$ and $\partial(\psi_2, \psi_3) / \partial(x^i, x^j) = 0$. Using these results, we calculate $(\partial \psi_2 / \partial x^s) \partial(\psi_1, \psi_3) / \partial(x^k, x^l) = (\partial \psi_2 / \partial x^s)(\partial \psi_1 / \partial x^k) \times (\partial \psi_3 / \partial x^l) - (\partial \psi_2 / \partial x^s)(\partial \psi_1 / \partial x^l)(\partial \psi_3 / \partial x^k) = (\partial \psi_1 / \partial x^k) \times (\partial \psi_2 / \partial x^s)(\partial \psi_3 / \partial x^l) - (\partial \psi_1 / \partial x^l)(\partial \psi_2 / \partial x^s)(\partial \psi_3 / \partial x^s) = (\partial \psi_3 / \partial x^s) \partial(\psi_1, \psi_2) / \partial(x^k, x^l) = 0$. But since for some value of s, $\partial \psi_2 / \partial x^s \neq 0$, we have $\partial(\psi_1, \psi_3) / \partial(x^k, x^l) = 0$, and hence $\Psi_1 \overset{T}{\approx} \Psi_3$.

Theorem III.22: Let $dq = A_k(x^i) dx^k$ be extensive on $\mathrm{Im}e(T)$. Integrable on an open connected subset $V \subset \mathrm{Im}e(T)$ with $A_k \neq 0$, $k = 1, 2, \ldots, n$, of class $C^{(1)}$ on V and $\partial A_k / \partial x^j - \partial A_j / \partial x^k \neq 0$ for some k, j. If dq admits an intensive integrating factor, it is unique to within a multiplicative constant.

Proof: Suppose M is an intensive integrating factor for dq and ϕ the corresponding pseudopotential. By Theorem III.12(4) any other integrating factor $\bar{M}$ can be expressed as $\bar{M} = Mg(\phi)$. Then, by differentiation, $\partial \bar{M} / \partial x^\alpha = g \partial M / \partial x^\alpha + M\dot{g} \partial \phi / \partial x^\alpha = g \partial M / \partial x^\alpha + \dot{g} M^2 A_\alpha$, and hence $x^\alpha \partial \bar{M} / \partial x^\alpha = gx^\alpha \partial M / \partial x^\alpha + \dot{g} M^2 x^\alpha A_\alpha$. If M and $\bar{M}$ are both intensive, then by Euler's theorem on homogeneous functions $x^\alpha \partial M / \partial x^\alpha = 0 = x^\alpha \partial \bar{M} / \partial x^\alpha$, and hence $\dot{g} M^2 x^\alpha A_\alpha = 0$. But $M \neq 0$, and by Theorem III.20 $x^\alpha A_\alpha \neq 0$ since $x^\alpha A_\alpha$ is not a constant function. Therefore, $\dot{g} = 0$ or $g = \mathrm{const}$.

This theorem establishes that all intensive integrating factors, if there are any, are linearly and thus functionally dependent. There only remains the task of showing that some empirical temperature is a member of the same equivalence class or else devising a set of criteria which could be used to test an empirical temperature.

Theorem III.23: Suppose $\Psi_1 \overset{T}{\approx} \Psi_2$, and the coordinate representation $\psi_1(x^i)$ of Ψ_1 is homogeneous of degree p_1 in the extensive variables, while the coordinate representation $\psi_2(x^i)$ of Ψ_2 is homogeneous of degree p_2 in the extensive variables. Then $p_1 = 0$ iff $p_2 = 0$.

Proof: Since $\Psi_1 \overset{T}{\approx} \Psi_2$, we have $G(\psi_1, \psi_2) = 0$ and $G[\psi_1(\lambda x^\alpha, x^\mu), \psi_2(\lambda x^\alpha, x^\mu)] = G[\lambda^{p_1} \psi_1(x^\alpha, x^\mu), \lambda^{p_2} \psi_2(x^\alpha, x^\mu)] = 0$. Differentiating with respect to λ and then setting $\lambda = 1$ gives $p_1 \psi_1 \partial G / \partial \psi_1 + p_2 \psi_2 \partial G / \partial \psi_2 = 0$. Since ψ_1, $\partial G / \partial \psi_1$, ψ_2, and $\partial G / \partial \psi_2$ are all nonzero, it follows that $p_1 = 0$ iff $p_2 = 0$.

Axiom III.24: Let Θ be an empirical temperature.

1604   J. Math. Phys., Vol. 17, No. 8, August 1976

Frank J. Zeleznik   1604

Then its coordinate representation $\theta(x^i)$ is intensive.

The combination of this axiom plus the previous theorem shows that the association of some empirical temperature with an integrating factor is possible only if $đq$ admits an intensive integrating factor (Theorem III.9 and Theorem III.20). But the existence of the essentially unique intensive integrating factor (Theorem III.22) does not establish its functional dependence with the coordinate representation, $\theta(x^i)$, of some empirical temperature. This must be proved either directly by implementing Definition III.1(2) or indirectly by using Corollary III.3. In either case one must know both the integrating factor and the empirical temperature. Is it possible to devise a test which involves only the empirical temperature? The answer is yes.

Theorem III.25: Let $đq = A_k(x^i)dx^k$ be extensive on $\mathrm{Im}e(T)$, integrable on an open, connected subset $V \subset \mathrm{Im}e(T)$ with $A_k \neq 0$, $k = 1, 2, \ldots, n$, of class $C^{(1)}$ on V, $\partial A_k/\partial x^j - \partial A_j/\partial x^k \neq 0$ for some k, j, and $x^\alpha A_\alpha \neq 0$. Let $\theta(x^i)$, the coordinate representation of an empirical temperature, be of class $C^{(1)}$ on V. Then ∃ an integrating factor $M = M(\theta)$ for $đq$ iff $(x^\alpha A_\alpha)^{-1}$ is not an integrating factor, and for some function $h(\theta)$, $\dot{h}(\theta) = dh/d\theta \neq 0$, the empirical temperature satisfies $(A_i \partial \theta/\partial x^j - A_j \partial \theta/\partial x^i)\dot{h}(\theta) = h(\theta)(\partial A_j/\partial x^i - \partial A_i/\partial x^j)$.

Proof: Suppose $M = M(\theta)$. Then $\partial M/\partial x^i = \dot{M}(\theta)\partial\theta/\partial x^i$, and substituting this into the differential equations determining M (Theorem III.9) gives the desired equation for θ. Conversely, if θ satisfies the given differential equations, then obviously $M = h(\theta)$ is an integrating factor. The balance of the theorem follows from the fact that any function of θ is intensive and Theorem III.20.

Suppose $\theta(x^i)$ is an empirical temperature based upon ideal gas thermometry (Adkins,[2] p. 20). Then thermodynamics assumes $M^{-1} = \theta + k$, where k is a real number and is 273.15 if the triple point of water is used as a fixed point. What are the consequences of the preceding theorem in this case?

Corollary III.26: Let $đq = A_k(x^i)dx^k$ be extensive on $\mathrm{Im}e(T)$, integrable on an open, connected subset $V \subset \mathrm{Im}e(T)$, with $A_k \neq 0$ of class $C^{(1)}$ on V, $\partial A_k/\partial x^j - \partial A_j/\partial x^k \neq 0$ for some k, j, and $x^\alpha A_\alpha \neq 0$. Let $\theta(x^i)$, the coordinate representation of an empirical temperature, be of class $C^{(1)}$ on V. Then ∃ an integrating factor for $đq$ of the form $M^{-1} = \theta + k > 0$, for some constant k, iff $(x^\alpha A_\alpha)^{-1}$ is not an integrating factor and $\theta + k$ satisfies $A_i \partial \ln(\theta + k)/\partial x^j - A_j \partial \ln(\theta + k)/\partial x^i = -(\partial A_j/\partial x^i - \partial A_i/\partial x^j)$.

Proof: Since $M^{-1} = \theta + k \Rightarrow \dot{M}^{-1} = -(\theta + k)^2$, then $\dot{M}/M = -(\theta + k)^{-1}$. Further, $(\theta + k)^{-1}\partial\theta/\partial x^i = \partial \ln(\theta + k)/\partial x^i$ and, now using Theorem III.25, the corollary follows.

While it is certainly possible that ideal gas thermometry generates a solution to the system of equations given in Corollary III.26, there is no *a priori* reason to expect that nature would be so cooperative. This is particularly true if n is at all large. It seems to me that a thermodynamicist would want to check on nature's beneficence.

Each of the topics discussed in this section has dealt with some aspect of the connection between entropy and the differential form of the quasistatic first law. Nowhere was it necessary either to invoke Caratheodory's theorem or to use composite systems. In all cases it was possible to phrase assumptions and results in terms of the functions $A_k(x^i)$ which determine $đq$. The advantages of such an approach are obvious since they expose the assumptions which must be made if thermodynamics, as presently practiced, is to be valid. It should be clear that the necessary assumptions are far from trivial.

IV. CONTINUUM CONSIDERATIONS

Every fundamental feature of thermodynamics has been carefully examined in the preceding three sections. The treatment was intentionally quite abstract, and no specification was ever made of the nature of measurements on X or of their restriction to Σ. Indeed, nothing was said about the sets X and Σ themselves. Neither was there any need to invoke a particular mechanics in the analysis. Hence, the mathematical results are independent of these factors. Applicability of some of the mathematics, of course, assumes that one can at least achieve a suitable realization of heat as well as an appropriate definition of the sets X and Σ. All of the theory can be used only if one can define a differential first law. The purpose of this concluding section is to give a realization of the mathematics based on the equations of continuum mechanics. That is, the equations of continuum mechanics will be used to derive a differential form of the first law. Now continuum mechanics describes a particular universe which is defined, though not necessarily uniquely, by a set of constitutive relations in the sense that the constitutive relations determine the system of continuum equations and the solutions of these equations represent the phenomena of that universe. The utility of the continuum approach is determined by one's ability to find some set of constitutive relations which defines a universe whose phenomena correspond to observations in the real world. At the present time one cannot point to a "best" set of constitutive relations, nor can one eliminate constitutive relations from continuum mechanics. This then implies that one should only deal in broad generalities when discussing constitutive relations. This approach offers one the advantage of not becoming mired in the details of the functional form for constitutive relations, and it also makes the results applicable for a wide class of constitutive relations.

Naturally the mathematics to be used in this section is completely different from the mathematics of the previous sections. Here we shall require tensor analysis on a three-dimensional Riemannian manifold and an adequate background is available from the book on tensor analysis written by Sokolnikoff.[22] Those who prefer a more abstract approach to the subject should consult the book by Bishop and Goldberg.[23] Notationally, Latin indices will represent tensor indices and will have the range 1, 2, 3, while Greek indices will be nontensor indices and will have an unspecified range. The spatial coordinates will be indicated by x^k, and t will represent the time. Covariant derivatives will be written as ∇_k, while absolute, or intrinsic, derivatives with respect to time will be written as $\delta/\delta t = \partial/\partial t + v^k \nabla_k$, where v^k is the velocity vector.

Axiom IV.1: To each measurement there corresponds a temporal evolution equation on a three dimensional Reimannian manifold.

The balance of the material in this section lends itself to a somewhat less formal approach than the axiom-theorem method of the previous sections. We shall need only a few of the evolution equations mentioned in Axiom IV.1 in order to obtain the first law in differential form. These are

$$\frac{\delta m}{\delta t} + m \nabla_k v^k = 0 = \frac{\partial m}{\partial t} + \nabla_k (m v^k), \qquad \text{(IV.2.1)}$$

$$\frac{\partial \mathcal{E}}{\partial t} + \nabla_k J_{\mathcal{E}}^k = 0, \qquad \text{(IV.2.2)}$$

$$m \frac{\delta n_\lambda}{\delta t} = R_\lambda - \nabla_k d_\lambda^k = \frac{\partial}{\partial t}(m n_\lambda) + \nabla_k (m n_\lambda v^k), \qquad \text{(IV.2.3)}$$

$$m \frac{\delta v_k}{\delta t} = F_k, \qquad \text{(IV.2.4)}$$

$$\frac{\delta e_{kl}}{\delta t} = B_{klij} g^{im} g^{jn} \nabla_n v_m, \qquad \text{(IV.2.5)}$$

$$\frac{\partial \rho}{\partial t} + \nabla_k j^k = 0, \qquad \text{(IV.2.6)}$$

$$\nabla_k E^k = \frac{\rho}{\epsilon_0}, \qquad \text{(IV.2.7)}$$

$$\nabla_l B^{kl} = \mu_0 \left(j^k + \epsilon_0 \frac{\partial E^k}{\partial t} \right), \qquad \text{(IV.2.8)}$$

$$\nabla_{[i} B_{jk]} \equiv \nabla_i B_{jk} + \nabla_j B_{ki} + \nabla_k B_{ij} = 0, \qquad \text{(IV.2.9)}$$

$$\nabla_k E_l - \nabla_l E_k = \frac{\partial B_{lk}}{\partial t}. \qquad \text{(IV.2.10)}$$

In these equations m is the mass density, v^k the velocity, $\mathcal{E}$ the energy density, $J_{\mathcal{E}}^k$ the flux of energy, n_λ the amount of species λ expressed per unit mass, R_λ the rate of production of species λ, d_λ^k the diffusive flux of species λ, F_k the force per unit volume, ϵ_{kl} the symmetric Eulerian strain tensor, B_{klij} a fourth rank tensor expressible in terms of e_{kl} and the metric tensor g_{ij}. The vector j^k is the flux of charge and is called the current density; ρ is the charge density, E_k the electric field, ϵ_0 the vacuum permittivity, B_{kl} the skew-symmetric magnetic induction tensor, and μ_0 the vacuum permeability.

Each of the first four equations is a conservation equation from continuum mechanics and, in order, they represent the conservation of mass, energy, species, and momentum. The last four equations are Maxwell's equations of electromagnetism. The evolution equation for the strain tensor (IV.2.5) follows by direct differentiation of the definition of the strain tensor

$$e_{kl}(x, t) \equiv \tfrac{1}{2}[g_{kl}(x) - G_{kl}(x, t)], \qquad \text{(IV.3.1)}$$

$$G_{kl}(x, t) \equiv \frac{\partial \bar{x}^i}{\partial x^k} \frac{\partial \bar{x}^j}{\partial x^l} g_{ij}(\bar{x}), \qquad \text{(IV.3.2)}$$

where G_{kl} is the symmetric Cauchy deformation tensor, $\bar{x}^i$ are the material or Lagrangian coordinates, and x^i are the spatial or Eulerian coordinates. The relation-ship between the two classes of coordinates describes the motion of the continuum

$$x^i = x^i(\bar{x}, t), \qquad \text{(IV.4.1)}$$

$$\bar{x}^i = \bar{x}^i(x, t), \qquad \text{(IV.4.2)}$$

with $x^i(\bar{x}, 0) = \bar{x}^i$. The tensor B_{klij} is given by

$$B_{klij} = \tfrac{1}{2}(g_{kl} G_{lj} + g_{li} G_{kj}). \qquad \text{(IV.5.1)}$$

The equation (IV.2.6) is the expression of charge conservation. Invariably the force F_k is broken down into conservative, nonconservative, and stress forces, while charge and current are partitioned into bound and free charges and currents. The following partitionings of F_k, ρ, and j_k are sufficiently general to accommodate the ordinary assumptions:

$$F^k = -m g^{ki} \nabla_i \Omega + f^k + \nabla_j \tau^{kj}, \qquad \text{(IV.6.1)}$$

$$\tau^{ij} = m S^{kl} B_{klmn} g^{mi} g^{nj} + \Delta \tau^{ij}, \qquad \text{(IV.6.2)}$$

$$\rho = \rho_b + \rho_f, \qquad \text{(IV.6.3)}$$

$$j^k = j_b^k + j_f^k, \qquad \text{(IV.6.4)}$$

$$\rho_b = -\nabla_k P^k - \mu_0 \epsilon_0 \frac{\partial \Phi}{\partial t}, \qquad \text{(IV.6.5)}$$

$$j_b^k = \frac{\partial P^k}{\partial t} + \nabla_j M^{kj} + g^{kl} \nabla_l \Phi, \qquad \text{(IV.6.6)}$$

$$D_k = \epsilon_0 E_k + P_k, \qquad \text{(IV.6.7)}$$

$$B_{ik} = \mu_0 (H_{ik} + M_{ik}). \qquad \text{(IV.6.8)}$$

In these equations Ω is the scalar force potential, f^k the nonconservative force, τ^{kj} the stress tensor, which need not be symmetric for our purposes, ρ_b and ρ_f the bound and free charges, and the bound and free currents j_b^k and j_f^k. The vector P^k is the polarization, the skew-symmetric tensor M^{kj} is the magnetization, D^k the electric displacement vector, and the skew-symmetric tensor H_{ik} is the magnetic field intensity. In the decomposition of the stress tensor given by (IV.6.2) the term $\Delta \tau^{ij}$ will shortly be seen to represent the dissipative stresses, while the first term, containing the tensor S^{kl}, will represent nondissipative stresses. Because of the symmetry $B_{klij} = B_{lkij}$ it is no restriction to assume that S^{kl} is symmetric. The terms involving the scalar Φ in (IV.6.5) and (IV.6.6) are new and the conventional case corresponds to $\Phi = 0$. The form of the added Φ terms was motivated by the desire to maintain the four-dimensional character of the electromagnetic equations. In the case $\Phi = 0$ the bound charge and bound current densities can be regarded as the components of a four-dimensional vector which can be expressed as the divergence of a four-dimensional, skew-symmetric, second rank tensor. The components of this tensor are expressible in terms of P^k and M^{kl}. The additional terms in (IV.6.5) and (IV.6.6), involving Φ, have been added so that in four-dimensional notation the bound charge-current 4-vector now appears as the divergence of the skew-symmetric tensor minus the four-dimensional gradient of Φ. The physical significance of the Φ terms can be most easily seen by deriving the equations satisfied by ρ_b, ρ_f, D^k, and H_{ik}. These are easily

obtained by combining Eqs. (IV.6.3)—(IV.6.8) with Eqs. (IV.2.6)—(IV.2.8):

$$\frac{\partial \rho_f}{\partial t} + \nabla_k j_f^k = \mu_0 \epsilon_0 \frac{\partial^2 \Phi}{\partial t^2} - \nabla^2 \Phi = -\left[\frac{\partial \rho_b}{\partial t} + \nabla_k j_b^k\right], \quad (IV.7.1)$$

$$\nabla_k D^k = \rho_f - \mu_0 \epsilon_0 \frac{\partial \Phi}{\partial t}, \quad (IV.7.2)$$

$$\nabla_l H^{kl} = j_f^k + \frac{\partial D^k}{\partial t} + g^{kl} \nabla_l \Phi. \quad (IV.7.3)$$

The Laplacian ∇^2 is defined to be $\nabla^2 \equiv g^{ij}\nabla_i\nabla_j$, and the d'Alembertian is defined to be $\Box = \mu_0\epsilon_0 \partial^2/\partial t^2 - \nabla^2$. Thus, the d'Alembertian of Φ, $\Box\Phi \equiv \mu_0\epsilon_0 \partial^2\Phi/\partial t^2 - \nabla^2\Phi$, acts as a source for free charge and a sink for bound charge and, in addition, the time derivative of Φ contributes a source term for the electric displacement vector, while its gradient acts as a source term for the magnetic field intensity. It should be clear that higher order tensors could be used to generate expressions for the force F_k in (IV.6.1) and for the bound charges and currents in (IV.6.5) and (IV.6.6); however, that will not be done here.

Related to the decomposition of F_k, ρ_b, and j_b^k is a similar decomposition of the energy flux $J_{\mathcal{E}}^k$. Rather than writing down the form immediately, as was done for F_k, ρ_b, and j_b^k, and then deducing its consequence, I shall proceed in reverse fashion. That is, I shall look at the consequences of the decomposition of F_k, ρ_b, and j_b^k and deduce from these an appropriate form for $J_{\mathcal{E}}^k$. Contracting the momentum Eq. (IV.2.4) with v^k, using the form (IV.6.1) for the force per unit volume and using the relation $\delta\Omega/\delta t = \partial\Omega/\partial t + v^k\nabla_k\Omega$ leads to the result

$$m \frac{\delta}{\delta t}\left(\frac{v^k v_k}{2} + \Omega\right) = m \frac{\partial \Omega}{\partial t} + v^k f_k + \nabla_j\left(\tau^{kj} v_k\right) - \tau^{kj} \nabla_j v_k. \quad (IV.8.1)$$

The term $v^k v_k/2$ is called the kinetic energy per unit mass and Ω is called the potential energy per unit mass. If the conservation of mass (continuity), Eq. (IV.2.1) is used to rewrite the left side of this equation, we get

$$\frac{\partial}{\partial t}\left[m\left(\frac{v^i v_i}{2} + \Omega\right)\right] + \nabla_k\left[m\left(\frac{v^i v_i}{2} + \Omega\right)v^k - \tau^{ik}v_i\right]$$
$$= m\frac{\partial \Omega}{\partial t} + v^k f_k - \tau^{kj}\nabla_j v_k. \quad (IV.8.2)$$

From two of Maxwell's equations [(IV.2.8), (IV.2.10)] it is possible to show that

$$\frac{\partial U}{\partial t} + \nabla_k N^k = -E_k j^k, \quad (IV.9.1)$$

$$U = \frac{\epsilon_0}{2}E^k E_k + \frac{1}{4\mu_0}B^{kl}B_{kl}, \quad (IV.9.2)$$

$$N^k = \frac{B^{kl}E_l}{\mu_0}, \quad (IV.9.3)$$

where U is called the energy density of the electromagnetic field and N^k is the Poynting vector. But the expressions (IV.6.5) and (IV.6.6) for ρ_b and j_b^k imply

$$E_k j_b^k = E_k \frac{\partial P^k}{\partial t} - \frac{1}{2}M^{kl}\frac{\partial B_{kl}}{\partial t} - \nabla_l(M^{lk}E_k - \Phi E^l) - \frac{\Phi \rho}{\epsilon_0}, \quad (IV.10.1)$$

$$\frac{\Phi \rho_b}{\epsilon_0} = -\nabla_k\left(\frac{\Phi P^k}{\epsilon_0}\right) + \frac{P^k}{\epsilon_0}\nabla_k \Phi - \frac{\mu_0}{2}\frac{\partial \Phi^2}{\partial t}, \quad (IV.10.2)$$

$$E_k j_b^k = E_k \frac{\partial P^k}{\partial t} - \frac{1}{2}M^{kl}\frac{\partial B_{kl}}{\partial t} - \nabla_l\left(M^{lk}E_k - \frac{\Phi D^l}{\epsilon_0}\right)$$
$$- \frac{\Phi \rho_f}{\epsilon_0} - \frac{P^k \nabla_k \Phi}{\epsilon_0} + \frac{\mu_0}{2}\frac{\partial \Phi^2}{\partial t}. \quad (IV.10.3)$$

To obtain the first of these equations, it was necessary to use the skew-symmetry of M^{ik} and Maxwell's equation (IV.2.10). The third member of this set is obtained by substituting (IV.10.2) into (IV.10.1) and using (IV.6.7). If $E_k j_b^k$ is eliminated from IV.9.1 by means of IV.10.3, we obtain

$$\frac{\partial U}{\partial t} + \nabla_l\left(H^{lk}E_k + \frac{\Phi D^l}{\epsilon_0}\right) = -E_k j_f^k + \frac{\Phi \rho_f}{\epsilon_0} + \frac{P^k \nabla_k \Phi}{\epsilon_2}$$
$$- \frac{\mu_0}{2}\frac{\partial \Phi^2}{\partial t} - E_k\frac{\partial P^k}{\partial t} + \frac{1}{2}M^{kl}\frac{\partial B_{kl}}{\partial t}. \quad (IV.11.1)$$

If the internal energy per unit mass u is defined by

$$mu = \mathcal{E} - U - m\left(\frac{v^k v_k}{2} + \Omega\right), \quad (IV.12.1)$$

then combining the evolution equations for $\mathcal{E}$ [(IV.2.2)], $v^k v_k/2 + \Omega$ [IV.8.2)], and U [IV.11.1)] produces an evolution equation for u in the form

$$\frac{\partial(mu)}{\partial t} + \nabla_l\left[-m\left(\frac{v^k v_k}{2} + \Omega\right)v^l + \tau^{il}v_i - H^{lk}E_k - \frac{\Phi D^l}{\epsilon_0} + J_{\mathcal{E}}^k\right]$$
$$= E_k j_f^k - \frac{\Phi \rho_f}{\epsilon_0} - \frac{P^k}{\epsilon_0}\nabla_k \Phi + \frac{\mu_0}{2}\frac{\partial \Phi^2}{\partial t} - m\frac{\partial \Omega}{\partial t} - v^k f_k$$
$$+ \tau^{kj}\nabla_j v_k + E_k\frac{\partial P^k}{\partial t} - \frac{1}{2}M^{kl}\frac{\partial B_{kl}}{\partial t}. \quad (IV.12.2)$$

Suppose we now require that the flux of u, J_u^k, have the form

$$J_u^k = muv^k + q^k + \mu^\lambda d_\lambda^k. \quad (IV.12.3)$$

The first term represents the convective flux of internal energy, q^k is the flux of internal energy in the absence of convection and diffusion of species, and the last term corresponds to the flux of internal energy due to the diffusion of species. A comparison of the flux J_u^k with the flux as given by the divergence term in (IV.12.2) leads to the identification

$$J_{\mathcal{E}}^l \equiv m\left(u + \frac{v^i v_i}{2} + \Omega\right)v^l + q^l + \mu^\lambda d_\lambda^l - \tau^{il}v_i + H^{lk}E_k + \frac{\Phi D^l}{\epsilon_0}$$

$$(IV.13.1)$$

and

$$m\frac{\delta u}{\delta t} = -\nabla_k\left(q^k + \mu^\lambda d_\lambda^k\right) + E_k j_f^k - \frac{\Phi \rho_f}{\epsilon_0} - \frac{P^k}{\epsilon_0}\nabla_k \Phi$$
$$+ \frac{\mu_0}{2}\frac{\partial \Phi^2}{\partial t} - m\frac{\partial \Omega}{\partial t} - v^k f_k + \tau^{kj}\nabla_j v_k$$
$$+ E_k\frac{\partial P^k}{\partial t} - \frac{1}{2}M^{kl}\frac{\partial B_{kl}}{\partial t} = \frac{\partial}{\partial t}(mu) + \nabla_k(muv^k). \quad (IV.13.2)$$

Equation (IV.13.2) is the last evolution equation which will be needed in obtaining the first law in differential form. Note that (IV.13.1) effectively defines the total flux of energy $J_{\mathcal{E}}^k$. The flux of energy associated with the stress, $\tau^{il}v_i$, can be rewritten in terms of the

tensors S^{kl} and $\Delta\tau^{kl}$. Similarly, the source term $\tau^{kj}\nabla_j v_k$ for the internal energy can be written in terms of S^{kl} and $\Delta\tau^{kl}$. The source term can be rewritten simply by using (IV.2.5) and (IV.6.2). To rewrite the flux term, it is necessary to use the symmetry of S^{kl} as well as the definition of B_{klmn} given in (IV.5.1). Thus,

$$(S^{kl}B_{klmn}g^{mi}g^{nj})v_i = \tfrac{1}{2}S^{kl}(g_{km}G_{ln}+g_{lm}G_{kn})g^{nj}v^m$$

$$= \tfrac{1}{2}S^{kl}(v_kG_{ln}+v_lG_{kn})g^{nj}=S^{kl}v_kG_{ln}g^{nj}$$

$$= v^kS_{kl}G^{lj}.$$

The symmetry of S^{kl} was used in the next to last step. Thus, we have

$$\tau^{ij}v_i = mv^kS_{kl}G^{lj}+\Delta\tau^{ij}v_i \qquad \text{(IV.14.1)}$$

and

$$\tau^{kj}\nabla_j v_k = mS^{kl}\frac{\delta e_{kl}}{\delta t}+\Delta\tau^{kj}\nabla_j v_k, \qquad \text{(IV.14.2)}$$

Given the evolution equations (IV.2.1)–(IV.2.10) and (IV.13.2), we still do not have a well-defined system of equations for we are still lacking the necessary constitutive relations. In these evolution equations the constitutive relations for R_λ, d_λ^k, Ω, f_k, S^{kl}, $\Delta\tau^{kl}$, P^k, Φ, M^{kl}, j_f^k, q^k, and μ^λ must be prescribed. These quantities are potentially functions of position x^k, time t, and all measurements, not only those whose evolution equations have been given, as well as their covariant and temporal derivatives. Fortunately for our purposes it is unnecessary to select a particular set of constitutive relations and we can proceed to the first law. Suppose we define a polarization per unit mass p^k and a magnetization per unit mass m^{kl} by

$$P^k = mp^k, \qquad \text{(IV.15.1)}$$

$$M^{kl} = mm^{kl}. \qquad \text{(IV.15.2)}$$

Then since $\delta P^k/\delta t = \partial P^k/\partial t + v^i\nabla_i P^k$ and $\delta B_{kl}/\delta t = \partial B_{kl}/\partial t + v^i\nabla_i B_{kl}$, it follows that

$$E_k\frac{\partial P^k}{\partial t}-\tfrac{1}{2}M^{kl}\frac{\partial B_{kl}}{\partial t}=mE_k\frac{\delta p^k}{\delta t}-E_k\nabla_i(P^k v^i)$$

$$-\tfrac{1}{2}mm^{kl}\frac{\delta B_{kl}}{\delta t}+\tfrac{1}{2}M^{kl}v^i\nabla_i B_{kl}. \qquad \text{(IV.15.3)}$$

If we use the evolution equation for species, (IV.2.3), we can write

$$\nabla_k(\mu^\lambda d_\lambda^k)=d_\lambda^k\nabla_k\mu^\lambda+\mu^\lambda\left(R_\lambda - m\frac{\delta n_\lambda}{\delta t}\right). \qquad \text{(IV.15.4)}$$

Now substituting (IV.14.2), (IV.15.3), and (IV.15.4) into the evolution equation for the internal energy, energy, (IV.13.2), we obtain

$$m\left[\frac{\delta u}{\delta t}-S^{kl}\frac{\delta e_{kl}}{\delta t}-E_k\frac{\delta p^k}{\delta t}+\tfrac{1}{2}m^{kl}\frac{\delta B_{kl}}{\delta t}-\mu^\lambda\frac{\delta n_\lambda}{\delta t}\right]$$

$$= -\nabla_k q^k+\Delta\tau^{kj}\nabla_j v_k-v^k f_k-m\frac{\partial\Omega}{\partial t}-\mu^\lambda R_\lambda-d_\lambda^k\nabla_k\mu^\lambda$$

$$+E_k j_f^k-\frac{\Phi\rho_f}{\epsilon_0}-\frac{P^k}{\epsilon_0}\nabla_k\Phi+\frac{\mu_0}{2}\frac{\partial\Phi^2}{\partial t}-E_k\nabla_i(P^k v^i)$$

$$+\tfrac{1}{2}M^{kl}v^i\nabla_i B_{kl}. \qquad \text{(IV.16.1)}$$

Based on this result we can now define heat and work

increments, thus giving explicit realizations for heat and work and hence connecting this section with the earlier ones:

$$\frac{\bar{\eth}Q}{\delta t} \equiv \frac{\delta u}{\delta t}-S^{kl}\frac{\delta e_{kl}}{\delta t}-E_k\frac{\delta p^k}{\delta t}+\tfrac{1}{2}m^{kl}\frac{\delta B_{kl}}{\delta t}-\mu^\lambda\frac{\delta n_\lambda}{\delta t},$$

$$\text{(IV.17.1)}$$

$$\frac{\bar{\eth}W}{\delta t} \equiv S^{kl}\frac{\delta e_{kl}}{\delta t}+E_k\frac{\delta p^k}{\delta t}-\tfrac{1}{2}m^{kl}\frac{\delta B_{kl}}{\delta t}+\mu^\lambda\frac{\delta n_\lambda}{\delta t},$$

$$\text{(IV.17.2)}$$

$$\bar{\eth}Q+\delta W=\delta u, \qquad \text{(IV.17.3)}$$

$$m\frac{\bar{\eth}Q}{\delta t}+\nabla_k q^k$$

$$= \Delta\tau^{kj}\nabla_j v_k-v^k f_k-m\frac{\partial\Omega}{\partial t}-\mu^\lambda R_\lambda-d_\lambda^k\nabla_k\mu^\lambda$$

$$+E_k j_f^k-\frac{\Phi\rho_f}{\epsilon_0}-\frac{P^k}{\epsilon_0}\nabla_k\Phi+\frac{\mu_0}{2}\frac{\partial\Phi^2}{\partial t}-E_k\nabla_i(P^k v^i)$$

$$+\tfrac{1}{2}M^{kl}v^i\nabla_i B_{kl}. \qquad \text{(IV.17.4)}$$

Here we have a realization of Axiom II.13, where differential heat and work were defined. It should be pointed out that the symbols $\bar{\eth}Q/\delta t$ and $\bar{\eth}W/\delta t$ are not to be interpreted as absolute derivatives. They are merely quantities defined by (IV.17.1) and (IV.17.2). If we were to commit this abuse of interpretation merely for descriptive purposes, then the equation for heat, (IV.17.4), suggests that q^k is the heat flux and that the right side of (IV.17.4) represents a source of heat, that is, dissipation. The dissipation contains contributions from, among others, stress $\Delta\tau^{kj}$, chemical reaction R_λ, diffusion d_λ^k, and free current j_f^k. The expressions for the heat and work increments show that heat and work are intimately connected with constitutive relations. Thus it is the constitutive relations for S^{kl}, p^k, m^{kl}, and μ^λ which preordain work in general and thermodynamic work in particular. Hence, making a statement about the integrability of the quasistatic limit, $\bar{\eth}q$, of $\bar{\eth}Q$ amounts to making a statement about the restrictions of S^{kl}, p^k, m^{kl}, and μ^λ to the thermodynamic subspace Σ. Except for p^k, these constitutive relations appear as coefficients of the incremental terms in (IV.17.1) and (IV.17.2). The stress—strain term, $S^{kl}\delta e_{kl}/\delta t$, in these equations is the appropriate form for solids, but it should reduce to pressure—volume work in the case of fluids. Suppose that the strain in a material can be characterized by a strain scalar e, that is, $e_{kl}=eg_{kl}$. Then from (IV.3.1) it follows that $G_{kl}=Gg_{kl}=(1-2e)g_{kl}$. Substituting these into the evolution equation for the strain tensor (IV.2.5) gives an evolution for the strain scalar $\delta e/\delta t = G\nabla_i v^i/3$. But if we use the continuity equation (IV.2.1) to eliminate $\nabla_i v^i$, we find $\delta e/\delta t = -(G/3m)\delta m/\delta t=(mG/3)$ $\times\delta(1/m)/\delta t$. Hence, $S^{kl}\delta e_{kl}/\delta t=S^{kl}g_{kl}\delta e/\delta t=(mGS_k^k/3)$ $\times\delta(1/m)/\delta t$. Since $1/m$ is the specific volume, it is reasonable to define the fluid pressure p as

$$-p \equiv mGS_k^k/3 \qquad \text{(IV.18.1)}$$

for then

$$S^{kl}\frac{\delta e_{kl}}{\delta t} = -p\frac{\delta(1/m)}{\delta t} \quad \text{if} \quad e_{kl}=eg_{kl}. \qquad \text{(IV.18.2)}$$

The differential equation $\delta e/\delta t = -(G/3m)\delta m/\delta t$
$= -[(1-2e)/3m]\delta m/\delta t$ can easily be integrated to give
$m(x,t) = m(\bar{x})(1-2e)^{3/2}$, where the initial state is taken
to be the zero strain state. This is nothing more than
a special case of the more general relation m
$= m(\bar{x})(|\delta_j^i - 2e_j^i|)^{1/2}$ where $|\delta_j^i - 2e_j^i|$ represents the
determinant of $\delta_j^i - 2e_j^i$.

Yet to be discussed are some points about the realiza-
tion of heat represented by (IV.17.1). First it should
be pointed out that, based on Definition III.13, u, S^{kl},
p^k, and m^{kl} are the obviously extensive quantities in $\eth Q$
while e_{kl}, E_k, B_{kl}, and μ^λ are intensive. For the fluid
case, pressure p is intensive and specific volume $1/m$
is extensive. Note that in $\eth Q$ not all of the differential
quantities are extensive. If the quasistatic limit of $\eth Q$,
that is, $\eth q$ is integrable and has an intensive integrating
factor in some subset of Σ, then this implies that the
Euler relation $\phi/M = x^\alpha A_\alpha$ (Theorem III.20) takes the
form $\phi/M = u - E_k p^k - \mu^\lambda n_\lambda$ for solids and $\phi/M = u + p/m$
$- E_k p^k - \mu^\lambda n_\lambda$ for fluids. Of course, all quantities on the
right side are meant to be the quasistatic forms.
Notable by their absence are the terms $S^{kl}e_{kl}$ and
$-1/2m^{kl}B_{kl}$. Another point to be mentioned concerns the
term $E_k \delta p^k$ in $\eth Q$. This tacitly assumes that in the
quasistatic case E_k must be regarded as a function of
p^k (and also u, e_{kl}, B_{kl}, n_λ). But conventionally the
polarization is regarded as a function of the electric
field. These two points of view are consistent iff the
Jacobian $|\partial P^k/\partial E^i|$ does not vanish in the quasistatic
limit and then $P^k = P^k(E^i, \ldots)$ can be inverted for E^i.
This is certainly satisfied when the polarization is a
linear function of the electric field, but a linear rela-
tionship is not generally applicable. If one would like to
use E_k, rather than p^k, as the independent variable,
then it is necessary to work with a Legendre trans-
form[11] of u. The appropriate function is $(u - p^k E_k)$. If
we use the conventional thermodynamic notation of s in
place of ϕ and T for M^{-1} then from Corollary III.11
we have, in the region of integrability,

$$\frac{\partial s}{\partial u} = T^{-1} = \left(\frac{\partial u}{\partial s}\right)^{-1},$$

$$T\frac{\partial s}{\partial e_{kl}} = -S^{kl} = -\frac{\partial u}{\partial e_{kl}} \quad \text{or} \quad T\frac{\partial s}{\partial(1/m)} = p = -\frac{\partial u}{\partial(1/m)},$$

$$T\frac{\partial s}{\partial p^k} = -E_k = -\frac{\partial u}{\partial p^k},$$

$$T\frac{\partial s}{\partial B_{kl}} = \frac{m^{kl}}{2} = -\frac{\partial u}{\partial B_{kl}},$$

$$T\frac{\partial s}{\partial n_\lambda} = -\mu^\lambda = -\frac{\partial u}{\partial n_\lambda}.$$

Here I have used the same symbol for the constitutive
quantity in Σ as in X. The partial derivatives of the
transform $u - p^k E_k$ are

$$\partial(u - p^k E_k)/\partial s = T,$$

$$\partial(u - p^k E_k)/e_{kl} = S^{kl} \quad \text{or} \quad \partial(u - p^k E_k)/\partial(1/m) = -p,$$

$$\partial(u - p^k E_k)/\partial E_k = -p^k,$$

$$\partial(u - p^k E_k)/\partial B_{kl} = -m^{kl}/2,$$

$$\partial(u - p^k E_k)/\partial n_\lambda = \mu^\lambda.$$

Similar expressions could be written for the Massieu
function which arises from a similar Legendre trans-
form of the entropy. Finally, Eq. (IV.17.4) can be
written in terms of s if the equation is restricted to the
region of integrability:

$$mT\frac{\delta s}{\delta t} + \nabla_k q^k$$

$$= \Delta T^{kj}\nabla_j v_k - v^k f_k - m\frac{\partial \Omega}{\partial t} - \mu^\lambda R_\lambda - d_\lambda^k \nabla_k \mu^\lambda + E_k j_f^k$$

$$- \frac{\Phi \rho_f}{\epsilon_0} - \frac{P^k}{\epsilon_0}\nabla_k \Phi + \frac{\mu_0}{2}\frac{\partial \Phi^2}{\partial t} - E_k \nabla_i(P^k v^i) + \tfrac{1}{2}M^{kl}v^i\nabla_i B_{kl}.$$

Hence, we now have an evolution equation for the
entropy, valid only in the region of integrability, and
its right side is interpretable as an entropy source. Of
course, since this equation is only valid in this region
of Σ, the quantities μ^λ, E_k, and M^{kl} could be replaced
by the appropriate partial derivatives of s.

CONCLUSIONS

Now that we have come to the end of the paper it is
appropriate to survey what has been achieved. First,
a complete examination of the fundamental mathematical
structure of thermodynamics has been made in order to
determine the function of each component of that struc-
ture. Second, nowhere were partitions, composite
systems, or the zeroth law of thermodynamics used in
the analysis. Finally, all assumptions were explicitly
stated in the development of the mathematical structure.

Out of this analysis a number of conclusions could be
drawn. The fundamental structure of thermodynamics is
completely algebraic. This structure requires the
notion of heat, but not the first law. It contains a pre-
cise definition of entropy which is identified as a purely
mathematical concept, and possesses criteria which,
if satisfied, are sufficient for the construction of an
entropy function from heat alone. Topological structure
serves only to permit a discussion of properties based
on the continuity of functions and the weak topology is
the appropriate topology for thermodynamics. If one
supposes the existence of a first law in differential
form, then one can discuss integrability of the first law
independently of Caratheodory's theorem and Caratheo-
dory's inaccessibility axiom. If the first law is inte-
grable, then the integrating factor may or may not be
a temperature, the pseudopotential may or may not be
an entropy and, indeed, it need not even be extensive.
Finally it is possible to construct a realization of the
first law from the equations of continuum mechanics
which is suitable for all systems whether they are solids
or fluids, whether they do or do not exhibit chemical
reactions and whether or not electromagnetic fields are
present.

What I have attempted to do is to elucidate the struc-
ture of thermodynamics and expose those assumptions
which must be satisfied if thermodynamics is to be
valid. Clearly some of the assumptions, such as Axiom
I.54, are well supported by experiment. Others, such
as the integrability of the quasistatic first law, though
regularly used are still unsubstantiated. Clearly the
unresolved questions should be experimentally
examined.

ACKNOWLEDGMENTS

[1]O. Redlich, Rev. Mod. Phys. **40**, 556 (1968).

[2]C. J. Adkins, *Equilibrium Thermodynamics* (McGraw-Hill, London, 1968).

[3]A. H. Wilson, *Thermodynamics and Statistical Mechanics* (Cambridge U. P., Cambridge, 1960).

[4]M. A. Zemansky, *Heat and Thermodynamics* (McGraw-Hill, New York, 1957), 4th ed.

[5]C. Caratheodory, Math. Ann. **67**, 355 (1909).

[6]P. T. Landsberg, *Thermodynamics* (Interscience, New York, 1961); Rev. Mod. Phys. **28**, 363 (1956).

[7]H. A. Buchdahl, Am. J. Phys. **17**, 41, 44, 212 (1949); **22**, 182 (1954); **23**, 65 (1955).

[8]B. Bernstein, J. Math. Phys. **1**, 222 (1960).

[9]J. B. Boyling, Commun. Math. Phys. **10**, 52 (1968); Proc. Roy. Soc. (Lond.) A **329**, 35 (1972).

[10]L. Tisza, *Generalized Thermodynamics* (M. I. T. Press, Cambridge, Mass., 1966).

[11]H. B. Callen, *Thermodynamics* (Wiley, New York, 1960).

[12]G. Falk and H. Jung, *Handbuch der Physik III/2* (Springer, Berlin, 1959), p. 119—75.

[13]R. Giles, *Mathematical Foundations of Thermodynamics* (Macmillan, New York, 1964).

[14]H. A. Buchdahl, Z. Phys. **152**, 425 (1958); **168**, 316 (1962).

[15]H. A. Buchdahl and W. Greve, Z. Phys. **168**, 386 (1962).

[16]P. Rastall, J. Math. Phys. **11**, 2955 (1970).

[17]J. L. B. Cooper, J. Math. Anal. Appl. **17**, 172 (1967).

[18]W. J. Hornix, in *A Critical Review of Thermodynamics*, edited by E. B. Stuart, B. Gal-Or, and A. J. Brainard (Mono, Baltimore, Md., 1970), pp. 235—52; Pure Appl. Chem. **22**, 237 (1970).

[19]H. Paley and P. M. Weichsel, *A First Course in Abstract Algebra* (Holt, Rinehart and Winston, New York, 1966).

[20]S. Willard, *General Topology* (Addison-Wesley, Reading, Mass., 1970).

[21]A. R. Forsyth, *Theory of Differential Equations* (Dover, New York, 1959), Vol. 1.

[22]I. S. Sokolnikoff, *Tensor Analysis* (Wiley, New York, 1964).

[23]R. L. Bishop and S. I. Goldberg, *Tensor Analysis on Manifolds* (Macmillan, New York, 1968).

Eigenvalues of the Casimir operators of the orthogonal and symplectic groups

C. O. Nwachuku* and M. A. Rashid[t]

International Center for Theoretical Physics, Trieste, Italy
(Received 30 September 1975)

Eigenvalues of the Casimir operators of the orthogonal and the symplectic groups are obtained in closed
and simple form by diagonalizing directly the matrices introduced by Perelomov and Popov. This method
unifies the treatment of the problem for the semisimple Lie groups.

I. INTRODUCTION

Many authors have attempted to compute the eigenvalues of the Casimir operators of the unitary, orthogonal, and symplectic groups which form the various series of the semisimple Lie groups.[1–8] Although, for the unitary groups, general expressions for the eigenvalues of these invariant operators of any degree are known in closed form,[9] no such expression seems to exist for the eigenvalues of the Casimir operators for the orthogonal and the symplectic groups. Wong and Yeh[7] have recently attempted to solve the problem, following Perelomov and Popov's work,[1,2] who obtained the eigenvalue of the invariant of the pth degree as sum of the elements of the pth power of a matrix which depends upon the group and the irreducible representation one considers. For the unitary groups, the corresponding matrix can very easily be diagonalized and the values of the invariants written down immediately in a closed and simple form.[9] On account of the relative complication of the corresponding matrices for the orthogonal and the symplectic groups, no one seems to have tried to diagonalize them. Wong and Yeh,[7] noting this difficulty, have attempted to circumvent it, which led them finally to quite complicated expressions for the eigenvalues.[10] We note that we do not require *all* the elements of the diagonalizing matrix or of its inverse, but only *sums* of elements in the rows of one and columns of the other which are nothing but left and right eigenvectors of these matrices.[11] In fact, most of the elements of the diagonalizing matrix are computable as products of certain factors, but some of the elements are single sums of products of factors. The sums which are required for computing the eigenvalues of the Casimir operators are indeed products of factors. This fact results in very simple answers for these values which closely resemble the answers for the unitary groups. Thus our method unifies the treatment of this problem for the various series of semisimple Lie groups.

We have organized this paper as follows. In Sec. II we obtain the matrices for the orthogonal and symplectic groups, which are required for our computation. In Sec. III we briefly sketch the diagonalization of these matrices and quote our results. Details of the diagonalization are given in Appendix A, whereas Appendix B contains the connection between left and right eigenvectors.

II. THE MATRICES A

We shall deal with the orthogonal and the symplectic groups together. Let $G^i{}_j$, where the indices i, j run from $-n$ to n for $O(2n + 1)$ and from $-n$ to n excluding zero for $O(2n)$ and $Sp(2n)$, be the infinitesimal generators of these groups. We then define the Casimir operator C_p of pth degree by

$$C_p = G^{i_1}{}_{i_2} G^{i_2}{}_{i_3} \cdots G^{i_p}{}_{i_1}, \tag{2.1}$$

where summation over repeated indices is understood. The generators $G^i{}_j$ satisfy the commutation relations

$$[G^i{}_j, G^k{}_l] = \delta^k{}_j G^i{}_l - \delta^i{}_l G^k{}_j + \epsilon_i \epsilon_j [\delta^{-i}{}_j G^k{}_{-i} - \delta^k{}_{-i} G^{-j}{}_l]. \tag{2.2}$$

The quantity ϵ_i in the above commutation relations is 1 for $O(N)$ (the symbol N will always mean $2n$ or $2n + 1$) whereas it takes the values $1, 0, -1$ for $i > 0$, $i = 0$, $i < 0$, respectively, for $Sp(2n)$. The above commutation relations can be obtained from the basic representation

$$G^i{}_j = x^i \frac{\partial}{\partial x^j} - \epsilon_i \epsilon_j x^{-j} \frac{\partial}{\partial x^{-i}}. \tag{2.3}$$

On account of (2.3), the generators $G^i{}_j$ have the symmetry

$$G^i{}_j = -\epsilon_i \epsilon_j G^{-j}{}_{-i}, \tag{2.4}$$

which results in $N(N - 1)/2$ and $n(n + 1)$ independent generators for $O(N)$ and $Sp(2n)$, respectively. In particular, for $O(N)$,

$$G^i{}_{-i} = 0.$$

The irreducible *tensor* representations of these groups are characterized by n integers $f_n, f_{n-1}, \ldots, f_1$, such that $f_n \geq f_{n-1} \geq \cdots \geq f_1$. These n integers are the eigenvalues of the n diagonal generators $G^i{}_i = -G^{-i}{}_{-i}$ ($i > 0$) in the highest state $| \phi \rangle$. Writing $f_{-i} = -f_i$, we generally have

$$G^i{}_i | \phi \rangle = f_i | \phi \rangle. \tag{2.5}$$

[Note that $G^0{}_0$ which can exist only for $O(2n + 1)$ is identically zero on account of the symmetry relation (2.4). This requires $f_0 = 0$ which is consistent with $f_{-i} = -f_i$.] The characterization of the *tensor* representations can also be made in terms of Young's tableau wherein the tensor representation $(f_n, f_{n-1}, \ldots, f_1)$ is shown as an n rowed diagram with f_i boxes in the ith row. Note that we are ordering our rows in the reverse direction. With this notation, the highest state is char-

acterized by filling the boxes in the ith row by the symbol i. The states of the irreducible representation represented by Young's tableau are ordered by comparing the ordered set $\{m_i: i=n,\ n-1,\ \ldots,1\}$ (called the weight of the state) of numbers of n's, $(n-1)$'s, etc., in the state, the higher state having a larger number of the *first* unequal number in the weights of the two states. In this sense, there is no state higher than the highest in the representation. The action of the generators A^i_j ($i>j$) on any state, therefore, is to transform it to a higher state. These generators must, therefore, annihilate the highest state $|\phi\rangle$. This argument will be made more precise later for any tensor operator.

The Casimir operators commute with all the generators and hence, by Schur's lemma are multiples of the identity matrix in any irreducible representation. Hence the eigenvalues of these invariants can be obtained by computing them on any state and in particular on the highest state. This is exactly what we shall do.

Let us also define

$$T^i_j(p) = \sum_{i_1,i_2,\ldots,i_{p-1}} G^i_{i_1}G^{i_1}_{i_2}\circ\circ\circ G^{i_{p-1}}_j, \qquad (2.6)$$

in terms of which

$$C_p = \sum_i T^i_i. \qquad (2.7)$$

The $\{T^i_j\}$ acts like a *tensor operator* whose commutation relations with the generators are given as

$$[G^i_j, T^k_l] = \delta^k_j T^i_l - \delta^i_l T^k_j + \epsilon_i\epsilon_j[\delta^{-i}_j T^k_{-l} - \delta^k_{-i}T^{-j}_l]. \qquad (2.8)$$

We show that

$$T^k_l|\phi\rangle = 0 \quad \text{for } k>l. \qquad (2.9)$$

Let $k>l \geq 0$. Then

$$[G^i_l, T^k_l] = (\delta^k_i - \delta^i_l + \delta^{-l}_i - \delta^k_{-i})T^k_l,$$

which shows that G^i_l commutes with T^k_l when $i>k$ but $[G^k_k,\ T^k_l] = T^k_l$. Thus $T^k_l|\phi\rangle$ has the weight $f_n, f_{n-1}, \ldots, f_{k+1}, f_k+1, \ldots$, which is higher than the weight of the highest state, which is impossible, or $T^k_l|\phi\rangle = 0$ where $k>l \geq 0$. If $l<0$, $k>l$ implies $-l>-k$ and the same argument shows that $T^k_l|\phi\rangle$ has weight $f_n, f_{n-1}, \ldots, f_{-l+1}, f_{-l}+1$ (c.: $f_{-l}+2$ if $k=-l$), $\ldots$, which is also higher than that of the highest state or again $T^k_l|\phi\rangle = 0$ where $l<0$.

Note that the property mentioned above is a property of any tensor operator satisfying the commutation relation (2.8) and does not depend upon whether it is constructed from the generators or not.

We can now write down a recursion scheme for computing $T^i_j(q)$ on the highest state. Indeed

$$T^i_j(q)|\phi\rangle = \sum_{j=-n}^{n} T^i_j(q-1)G^j_i|\phi\rangle$$

$$= \sum_{j\leq i} T^i_j(q-1)G^j_i|\phi\rangle \quad (\because G^j_i|\phi\rangle = 0 \text{ for } j>i)$$

$$= f_i T^i_i(q-1)|\phi\rangle + \sum_{j<i} T^i_j(q-1)G^j_i|\phi\rangle$$

[from (2.5)]

$$= f_i T^i_i(q-1)|\phi\rangle - \sum_{j<i} [G^j_i, T^i_j(q-)]|\phi\rangle$$

$$(\because T^i_j(q-1)|\phi\rangle = 0 \text{ for } i>j)$$

$$= f_i T^i_i(q-1)|\phi\rangle - \sum_{j<i} (1\mp\delta^{-j}_i)[T^j_j(q-1) - T^i_i(q-1)]|\phi\rangle$$

$$= \sum_{j=-n}^{n} A_{ij}T^j_j(q-1)|\phi\rangle, \qquad (2.10)$$

where

$$A_{ij} = [f_i + n + i - \tfrac{1}{2}(1+e_i)]\delta_{ij} - \theta_{ji}$$
$$+ \tfrac{1}{2}(1+e_i)\delta_{i,-j} \quad \text{for } O(2n+1)$$

$$= (f_i + n + i - 1 - e_i)\delta_{ij} - \theta_{ji} + \tfrac{1}{2}(1+e_i)\delta_{i,-j} \quad \text{for } O(2n)$$

$$= (f_i + n + i)\delta_{ij} - \theta_{ji} - \tfrac{1}{2}(1+e_i)\delta_{i,-j} \quad \text{for } Sp(2n). \qquad (2.11)$$

The symbol e_i in the above equation is $1, 0, -1$ for $i>0$, $i=0$, $i<0$, respectively, and $\theta_{ij} = 0,1$ for $i\geq j$, $i<j$, respectively. Note that $\mp$ in Eq. (2.10) are for $O(N)$ and $Sp(2n)$, respectively. The matrices A with the components A_{ij} given in (2.11) are those we desired to compute. These were first given by Perelomov and Popov.[1,2,12] Each one is a *fixed* (independent of q) $m \times m$ matrix which depends only upon the representation of the group (m is the order of the group).

The recursion relation (2.10) can now be repeatedly used taking into account the fixed nature of the matrix A to arrive at

$$T^i_j(p)|\phi\rangle = \sum_{j=-n}^{n} (A^p)_{ij}|\phi\rangle, \qquad (2.12)$$

or

$$C_p|\phi\rangle = \sum_i T^i_i(p)|\phi\rangle = \sum_{i,j=-n}^{n} (A^p)_{ij}|\phi\rangle.$$

Thus

$$C_p = \sum_{i,j=-n}^{n} (A^p)_{ij}. \qquad (2.13)$$

The matrices A in (2.11) are upper triangular and hence their eigenvalues λ_i ($-n \leq i \leq n$) are the diagonal elements. These can be written down immediately from (2.11) and are given below with respect to our ordering.

Group	Eigenvalues of the matrix A		
$O(2n+1)$:	$f_i + n + i - 1$ for $1 \leq i \leq n$,	n for $i=1$,	$f_i + n + i$ for $-n \leq i \leq -1$,
$O(2n)$:	$f_i + n + i - 2$ for $1 \leq i \leq n$,	$+f_i + n + i$ for $-n \leq i \leq -1$,	
$Sp(2n)$	$f_i + n + i$ for $1 \leq i \leq n$,	$+f_i + n + i$ for $-n \leq i \leq -1$.	

$$(2.14)$$

C.O. Nwachuku and M.A. Rashid

Since $f_n \geq f_{n-1} \geq \cdots \geq f_1$, all the eigenvalues of Sp(2n) are distinct and *strictly* decreasing. For $O(N)$, these also have the same property in general. However, when $f_1 = 0$ (the lowest row having no blocks) for $O(2n+1)$, the eigenvalues corresponding to $i = 1, 0$ are both n, whereas, for $O(2n)$, the eigenvalues corresponding to $i = \pm 1$ are both $n - 1$. Except for this case, the eigenvalues are all distinct and strictly decreasing. This remark will be useful when we attempt to diagonalize the nonsymmetric matrices A for the various types of groups.

III. DIAGONALIZATION OF THE MATRICES A FOR THE VARIOUS GROUPS

We have already remarked that the matrices A are upper triangular and we know their eigenvalues, which are *in general* distinct. Thus, to diagonalize, we must obtain the right and left eigenvectors X_α, Y_β (X_α is a column vector whereas Y_β is a row vector) with components $X_{i\alpha}$, $Y_{\beta j}$, respectively, forming the matrices X and Y. {The Greek indices also have ranges $-n$ to n [excluding zero for $O(2n)$ and Sp(2n)] but we use them to enumerate the eigenvectors.} The matrices X and Y are also upper triangular, and we choose their diagonal elements as unity. The eigenvectors are now completely defined though *not* normalized to unity. However, this choice has two important advantages. First, as will be shown in Appendix B, the matrix Y can be trivially written down from X. As a consequence, X and Y look similar in contrast to the choice of the diagonalizing matrix by Okubo[6] and Wong and Yeh.[7] Secondly, this choice automatically makes $Y = X^{-1}$ and diagonalization is achieved. Details of the diagonalization are given in Appendices A and B. In Appendix A we compute X whereas in Appendix B we show that Y can be obtained from X quite easily. We remark that, since A is nonsymmetric, Y is *not* the transpose of A as is also evident from the fact that both X and Y are upper triangular.

Noting now that $Y = X^{-1}$, we have

$$A_D = YAX \Rightarrow A = XA_D Y \Rightarrow A^p = XA^p_D Y,$$

or

$$\sum_{i,j} (A^p)_{ij} = \sum_{i,j,\alpha,\beta} X_{i\alpha}(A^p_D)_{\alpha\beta} Y_{\beta j},$$

where

$$(A^p_D)_{\alpha\beta} = \lambda^p_\alpha \delta_{\alpha\beta}.$$

Thus

$$C_p = \sum_{i,j}(A^p)_{ij} = \sum_\alpha \lambda^p_\alpha \sum_j Y_{\alpha j} \sum_i X_{i\alpha}, \qquad (3.1)$$

i.e., to compute C_p we require *only the products of sums of components of the left and right eigenvectors of the matrices A for the same eigenvalue λ_α*. This remark makes the computation of C_p a much simpler task.

The sums $\sum_i X_{i\alpha}$ are computed in Appendix A and to obtain $\sum_j Y_{\alpha j}$ we must use the rule in Appendix B. We quote below these results:

$$\sum_i X_{i\alpha} = \prod_{j=\alpha+1}^{n}(1 + \lambda_j - \lambda_\alpha)\Big/\prod_{j=\alpha+1}^{n}(\lambda_j - \lambda_\alpha) \quad (\alpha \geq 0) \qquad (3.2)$$

$$= \left\{\begin{array}{l} (1 + \lambda_0 - \lambda_\alpha)(1 + \lambda_{-\alpha} - \lambda_\alpha) - (\lambda_0 - \lambda_\alpha) \\ \lambda_{-\alpha} - \lambda_\alpha \\ 2 + \lambda_{-\alpha} - \lambda_\alpha \end{array}\right\}$$

$$\times \left(\prod_{j=\alpha+1}^{n}{}'(1 + \lambda_j - \lambda_\alpha)\Big/\prod_{j=\alpha+1}^{n}(\lambda_j - \lambda_\alpha)\right) \quad (\alpha \leq 0), \qquad (3.3)$$

$$\sum_i Y_{\alpha i} = \left\{\begin{array}{l} (1 + \lambda_0 - \lambda_\alpha)(1 + \lambda_{-\alpha} - \lambda_\alpha) - (\lambda_0 - \lambda_\alpha) \\ \lambda_{-\alpha} - \lambda_\alpha \\ 2 + \lambda_{-\alpha} - \lambda_\alpha \end{array}\right\}$$

$$\times \left(\prod_{j=-n}^{\alpha-1}{}'(1 + \lambda_j - \lambda_\alpha)\Big/\prod_{j=-n}^{\alpha-1}(\lambda_j - \lambda_\alpha)\right) \quad (\alpha \geq 0) \qquad (3.4)$$

$$= \prod_{j=-n}^{\alpha-1}(1 + \lambda_j - \lambda_\alpha)\Big/\prod_{j=-n}^{\alpha-1}(\lambda_j - \lambda_\alpha) \quad (\alpha \leq 0), \qquad (3.5)$$

where the primes on the products in (3.3) and (3.4) imply that $j = 0$, $-\alpha$ are to be omitted and the outside factor in these equations is for $O(2n+1)$, $O(2n)$, and Sp(2n), respectively.

Substituting from the above Eqs. (3.2)—(3.5) into (3.1), we finally arrive at

$$C_p = \sum_{\alpha=-n}^{n} \lambda^p_\alpha \left\{\begin{array}{l} (1 + \lambda_0 - \lambda_\alpha)(1 + \lambda_{-\alpha} - \lambda_\alpha) - (\lambda_0 - \lambda_\alpha) \\ \lambda_{-\alpha} - \lambda_\alpha \\ 2 + \lambda_{-\alpha} - \lambda_\alpha \end{array}\right\}$$

$$\times \left(\prod_{\substack{i=-n \\ i\neq 0,\pm\alpha}}^{n}(1 + \lambda_i - \lambda_\alpha)\Big/\prod_{\substack{i=-n \\ i\neq\alpha}}^{n}(\lambda_i - \lambda_\alpha)\right). \qquad (3.6)$$

Naturally for $O(2n)$ and Sp(2n) the product in the denominator in (3.6) should not contain $i = 0$. The same remark applies to the summation.

We have always been mentioning the tensor representations. Spinor representations exist for $O(2n)$ groups. In fact they appear as pairs of inequivalent representations with the same dimensions and the same values for all the Casimir operators we have defined. For a complete characterization in this case, we require another Casimir operator in place of C_{2n} in the independent set $C_2, C_4, \ldots, C_{2n}$. This operator C'_n and its eigenvalue are both given in Ref. 2.

Our expression (3.5) has the form of the corresponding expression[13]

$$C_p = \sum_\alpha \lambda^p_\alpha \times \left(\prod_{\substack{i=1 \\ i\neq\alpha}}^{n}(1 + \lambda_i - \lambda_\alpha)\Big/\prod_{\substack{i=1 \\ i\neq\alpha}}^{n}(\lambda_i - \lambda_\alpha)\right), \qquad (3.7)$$

for the unitary groups. Our method has thus unified the treatment of this problem for the various series of the semisimple Lie groups.

ACKNOWLEDGMENTS

1613   J. Math. Phys., Vol. 17, No. 8, August 1976

C.O. Nwachuku and M.A. Rashid   1613

to the Swedish International Development Authority for making possible their Associateship at the ICTP.

APPENDIX A: COMPUTATION OF THE RIGHT EIGENVECTORS OF THE MATRICES A

We first discuss the groups $O(2n)$ and $Sp(2n)$. We order the rows and the columns by the index values n, $n-1, \ldots, -n$. (The index 0 is missing in this case.) The matrix A is upper triangular having the diagonal elements λ_α $(-n \leq \alpha \leq n)$ given in (2.14) and has -1 in all the off-diagonal positions except the $(i, -i)$ positions for $1 \leq i \leq n$, where it has 0 or -2 for $O(2n)$ and $Sp(2n)$ cases, respectively. For the right eigenvector X_α corresponding to the eigenvalue λ_α, we immediately have $X_{i\alpha} = 0$ for $-n \leq i \leq \alpha - 1$. We shall attempt to compute $X_{i\alpha}$ for $\alpha + 1 \leq i \leq n$ in terms of $X_{\alpha\alpha}$, which we shall finally put equal to unity.

For $\alpha \geq 1$, the computation is trivial. In fact, the presence of off-diagonal elements not equal to -1 has no effect on this computation since $X_{i\alpha} = 0$ for $-n \leq i \leq \alpha - 1$. For this case, one finds immediately

$$X_{i\alpha} = X_{\alpha\alpha} \times \left(\prod_{j=\alpha+1}^{i-1} (1 + \lambda_j - \lambda_\alpha) \bigg/ \prod_{j=\alpha+1}^{i} (\lambda_j - \lambda_\alpha) \right) (\alpha + , \leq i \leq n)$$

(A1)

for the nonzero components, and

$$\sum_{i=-n}^{n} X_{i\alpha} = \sum_{i=\alpha}^{n} X_{i\alpha}$$

$$= X_{\alpha\alpha} \times \left(\prod_{j=\alpha+1}^{n} (1 + \lambda_j - \lambda_\alpha) \bigg/ \prod_{j=\alpha+1}^{n} (\lambda_j - \lambda_\alpha) \right)$$

(A2)

for the sum of the nonzero components.

Next we consider the somewhat more difficult case of $\alpha \leq -1$. The components $X_{i\alpha}$ $(\alpha + 1 \leq i \leq -1)$ are easily computable in terms of $X_{\alpha\alpha}$ to give

$$X_{i\alpha} = X_{\alpha\alpha} \times \left(\prod_{j=\alpha+1}^{i-1} (1 + \lambda_j - \lambda_\alpha) \bigg/ \prod_{j=\alpha+1}^{i} (\lambda_j - \lambda_\alpha) \right)$$

$$(\alpha + 1 \leq i \leq -1).$$

(A3)

In addition, the components $X_{i\alpha}$ $(-\alpha + 2 \leq i \leq n)$ are also computable but in terms of $X_{-\alpha+1,\alpha}$ in the form

$$X_{i\alpha} = X_{-\alpha+1\,\alpha} \times \left(\prod_{j=-\alpha+1}^{i} (1 + \lambda_j - \lambda_\alpha) \bigg/ \prod_{j=-\alpha+2}^{i-1} (\lambda_j - \lambda_\alpha) \right)$$

$$(-\alpha + 2 \leq i \leq n).$$

(A4)

In particular

$$X_{n\alpha} = X_{-\alpha+1\,\alpha} \times \left(\prod_{j=-\alpha+1}^{n-1} (1 + \lambda_j - \lambda_\alpha) \bigg/ \prod_{j=-\alpha+2}^{n} (\lambda_j - \lambda_\alpha) \right).$$

(A5)

Now the equation corresponding to the component $X_{n\alpha}$ is

$$(\lambda_n - \lambda_\alpha) X_{n\alpha} = \sum_{i=\alpha}^{n-1} X_{i\alpha},$$

which gives

$$\sum_{i=\alpha}^{n} X_{i\alpha} = (1 + \lambda_n - \lambda_\alpha) X_{n\alpha}.$$

By using (A5), this becomes

$$\sum_{i=-n}^{n} X_{i\alpha} = \sum_{i=\alpha}^{n} X_{i\alpha} = X_{-\alpha+1\,\alpha} \times \left(\prod_{j=-\alpha+1}^{n} (1 + \lambda_j - \lambda_\alpha) \bigg/ \prod_{j=-\alpha+2}^{n} (\lambda_j - \lambda_\alpha) \right).$$

(A6)

The above equation shows that we need to express $X_{-\alpha+1,\alpha}$ only in terms of $X_{\alpha\alpha}$ to be able to compute the sum of nonnegative components of the eigenvector X_α $(\alpha \leq -1)$, though the computation of the diagonalizing matrix will also require the components $X_{i\alpha}$ $(1 \leq i \leq -\alpha + 1)$ in terms of $X_{\alpha\alpha}$ since all other components are already given in (A3) and (A4). To compute these, we write the equations corresponding to the components $X_{i\alpha}$, $1 \leq i \leq -\alpha + 1$, which simplify to the form

$$(\lambda_{-\alpha+1} - \lambda_\alpha) X_{-\alpha+1\,\alpha} - (1 + \lambda_{-\alpha} - \lambda_\alpha) X_{-\alpha\alpha} = \pm X_{\alpha\alpha},$$

$$(\lambda_{-\alpha} - \lambda_\alpha) X_{-\alpha\alpha} - (1 + \lambda_{-\alpha-1} - \lambda_\alpha) X_{-\alpha-1\,\alpha} = \pm (X_{\alpha+1\,\alpha} - X_{\alpha\alpha}),$$

$$\cdots$$

$$(\lambda_2 - \lambda_\alpha) X_{2\alpha} - (1 + \lambda_1 - \lambda_\alpha) X_{1\alpha} = \pm (X_{-1\alpha} - X_{-2\alpha}),$$

$$(\lambda_1 - \lambda_\alpha) X_{1\alpha} = \begin{Bmatrix} 0 \\ 2 \end{Bmatrix} X_{-1\alpha} + \sum_{j=\alpha}^{-2} X_{j\alpha}.$$

(A7)

The above is a system of *inhomogeneous* equations with the right-hand sides known on account of (A3) in terms of

$$\sum_{j=\alpha}^{-2} X_{j\alpha} = (\lambda_{-1} - \lambda_\alpha) X_{-1\alpha} = \prod_{j=\alpha+1}^{-2} (1 + \lambda_j - \lambda_\alpha) \bigg/ \prod_{j=\alpha+1}^{-2} (\lambda_j - \lambda_\alpha).$$

The $\pm$ and $\begin{Bmatrix} 0 \\ 2 \end{Bmatrix}$ appearing on the right-hand sides are for $O(2n)$ and $Sp(2n)$, respectively.

From these equations we can solve for any of the $X_{i\alpha}$ $(1 \leq i \leq -\alpha + 1)$ in terms of $X_{\alpha\alpha}$ as the ratio of two determinants. The common determinant in the denominator can be calculated. It is just $\prod_{j=1}^{-\alpha+1} (\lambda_j - \lambda_\alpha)$. The determinants in the numerator can be expanded in terms of the known right-hand sides (which are products of factors) as a single sum. For our problem, we are only interested in $X_{-\alpha+1,\alpha}$ which appears in (A6). In the corresponding determinant, the terms can *fortunately* be summed to give

$$X_{-\alpha+1\,\alpha} = \begin{Bmatrix} \lambda_{-\alpha} - \lambda_\alpha \\ 2 + \lambda_{-\alpha} - \lambda_\alpha \end{Bmatrix} \times X_{\alpha\alpha}$$

$$\times \left(\prod_{j=\alpha+1}^{-\alpha-1} (1 + \lambda_j - \lambda_\alpha) \bigg/ \prod_{j=\alpha+1}^{-\alpha+1} (\lambda_j - \lambda_\alpha) \right).$$

(A8)

Now, using (A6), we finally arrive at

$$\sum_{i=-n}^{n} X_{i\alpha} = \sum_{i=\alpha}^{n} X_{i\alpha} = \begin{Bmatrix} \lambda_{-\alpha} - \lambda_\alpha \\ 2 + \lambda_{-\alpha} - \lambda_\alpha \end{Bmatrix} \times X_{\alpha\alpha}$$

$$\times \left(\prod_{j=\alpha+1}^{n}{}' (1 + \lambda_j - \lambda_\alpha) \bigg/ \prod_{j=\alpha+1}^{n}{}' (\lambda_j - \lambda_\alpha) \right),$$

(A9)

where the prime on the product symbol indicates that the term with $j = -\alpha$ is to be omitted. Note that none of the symbols in the above sum is to take the value zero. The method of calculation we have adopted works when $\alpha < -1$ and for $\alpha = -1$, $\alpha + 1$ does take the value

C.O. Nwachuku and M.A. Rashid   1614

zero. But, if we computed directly the case of $\alpha = -1$, the answer is exactly as in (A9) by omitting any term in which the index $j = 0$.

We now return to the problem of the equality of two eigenvalues for the groups $O(2n)$ which occurs when $f_1 = 0$, in which case $\lambda_1 = \lambda_{-1} = n - 1$. Let us look at the sum in (A9) which is valid whenever $f_1 \neq 0$ or $\lambda_1 - \lambda_{-1} \neq 0$ even for $\alpha = -1$. In this case, the product in the denominator starts at $j = -1$, i.e., with the factor $\lambda_1 - \lambda_{-1}$ which vanishes when $f_1 = 0$. This factor is precisely killed by the factor $(\lambda_1 - \lambda_{-1})$ appearing as a separate factor in (A9) in the numerator [since we are finally going to put $X_{\alpha\alpha} = -1$ for all, we notice that when $\lambda_1 - \lambda_{-1} = 0$ $(f_1 = 0)$, the sums in (A2) and (A11) are equal]. To be more precise, consider the case where $\alpha = -1$ and $\lambda_1 \neq \lambda_{-1}$. Then, for $\alpha = -1$, $X_{-2,-1} = X_{-3,-1} \cdots = X_{-n,-1} = 0$ and $X_{1,-1} = 0$, while, for $\alpha = +1$, $X_{-1,1} = \cdots = X_{-n,1} = 0$ and the calculation for the remaining components proceeds exactly as before. Taking $X_{11} = X_{-1,-1} = 0$, we see that the components $2 \leq i \leq n$ are exactly the same, and this explains the equality of the sums when $\lambda_1 = \lambda_{-1}$. Had we done the calculation starting from $\lambda_1 = \lambda_{-1}$ in the beginning, we would have found that the ± 1 components are arbitrary. Once one chooses for one of the vectors the $+1$ component zero and -1 component 1 and makes the opposite choice for the second vector, one arrives at what we obtained before. In this case, the formal and the exact methods bring about the same result. This favorable situation is the result of the fact that the matrix A is *nondefective*.

Finally, we examine the case of $O(2n + 1)$. In complete analogy with the $O(2n)$ case, the vectors with $\alpha \geq 0$ can be computed and these are given by (A1) and (A2) with $\alpha = 0$. When $\alpha < 0$, only the last of the equations in (A7) needs a change, and we must now add $X_{\alpha\alpha}$ also to the right-hand side, which is also given in (A3) for $i = 0$. The remaining steps are identical, and we finally arrive at the modifications of (A8) and (A9) as follows:

$$X_{-\alpha+1\,\alpha} = [(1 + \lambda_0 - \lambda_\alpha)(1 + \lambda_{-\alpha} - \lambda_\alpha) - (\lambda_0 - \lambda_\alpha)]$$
$$\times X_{\alpha\alpha} \times \left(\prod_{\substack{j=\alpha+1 \\ j\neq 0}}^{-\alpha-1} (1 + \lambda_j - \lambda_\alpha) \prod_{j=\alpha+1}^{-\alpha+1} (\lambda_j - \lambda_\alpha) \right) \quad \text{(A10)}$$

and

$$\sum_{i=-n}^{n} X_{i\alpha} = \sum_{i=\alpha}^{n} X_{i\alpha} = [(1 + \lambda_0 - \lambda_\alpha)(1 + \lambda_{-\alpha} - \lambda_\alpha) - (\lambda_0 - \lambda_\alpha)]$$
$$\times X_{\alpha\alpha} \times \left(\prod_{j=\alpha+1}^{n}{}' (1 + \lambda_j - \lambda_\alpha) \Big/ \prod_{j=\alpha+1}^{n} (\lambda_j - \lambda_\alpha) \right), \text{(A11)}$$

where the prime on the product indicates that $j = 0$, $-\alpha$ are to be omitted. Note that though we computed the answer taking $\alpha < 0$, the answer is correct even for $\alpha = 0$ when it coincides precisely with (A2) about which we have already remarked that it holds for $O(2n + 1)$ for all $0 \leq \alpha \leq n$.

We return again to the problem of the eigenvalues λ_0 and λ_1 becoming equal when $f_1 = 0$. In this case, if we solve the equations, only one independent vector is obtained, which is exactly the same as given in (A1) for $\alpha = 1$ and the sum of its components in (A2) is a multiple of X_{11}. However, if we look at the sum in (A2) for $\alpha = 0$

with $\lambda_1 = \lambda_0$ we see that the denominator has a factor $\lambda_1 - \lambda_0$ which vanishes in this case. *This is indicative of the defect of the matxix, and we cannot, in principle, diagonalize the matrix in this case.* To complete our computation even for this case, we proceed, formally supposing, in the beginning, that $\lambda_1 - \lambda_0 \neq 0$, and examine the left eigenvector also. We can immediately solve the equations and obtain the sum of the components of the left eigenvector as

$$Y_{\alpha\alpha} \times \left(\prod_{j=-n}^{-\alpha-1} (1 + \lambda_j - \lambda_\alpha) \Big/ \prod_{j=-n}^{-\alpha-1} (\lambda_j - \lambda_\alpha) \right)$$

for $\alpha \leq 0$. For $\alpha = 0$, we note that the numerator has a factor $1 + \lambda_{-1} - \lambda_0$. Since, for $O(2n + 1)$, $\lambda_{-i} + \lambda_i = \lambda_{-i} + \lambda_i = 2n - 1$ (for $i \neq 0$) and $\lambda_0 = n$, $1 + \lambda_{-1} - \lambda_0 = \lambda_0 - \lambda_1$, i.e., this factor is exactly the same (but opposite in sign) as the singular factor in the sum of the components of the right eigenvector. The product of these two, which is what we require, does not contain any singular character even for $\lambda_1 = \lambda_0$ $(f_1 = 0)$.

APPENDIX B: CALCULATION OF THE LEFT EIGENVECTORS

Note that the matrix A for each group is not symmetric. Thus, *a priori*, we do not expect any connection between the left and the right eigenvector. However, these matrices satisfy a pseudosymmetry given by

$$A_{ij}(\lambda_\alpha) = A_{-j,-i}(\lambda_{-\alpha}), \quad \text{(B1)}$$

which is the symmetry with respect to reflection in the diagonal joining $(n, -n)$ and $(-n, n)$ elements. This reflection only changes the diagonal elements and this change is incorporated in (B1).

Let us now examine the eigenvalue problem

$$A(\lambda_\beta)X_\alpha(\lambda_\beta) = \lambda_\alpha X_\alpha(\lambda_\beta), \quad \text{(B2)}$$

where X_α is a right eigenvector corresponding to the eigenvalue λ_α. Its components depend upon the eigenvalues of the matrix A and their ordering.

Writing out (B2) completely gives

$$\sum_j A_{ij}(\lambda_\beta)X_{j\alpha}(\lambda_\beta) = \lambda_\alpha X_{i\alpha}(\lambda_\beta). \quad \text{(B3)}$$

By using (B1), this becomes

$$\sum_j X_{j\alpha}(\lambda_\beta)A_{-j,-i}(\lambda_{-\beta}) = \lambda_\alpha X_{i\alpha}(\lambda_\beta). \quad \text{(B4)}$$

Let us write

$$Y_{-\alpha,-i}(\lambda_{-\beta}) = X_{i\alpha}(\lambda_\beta). \quad \text{(B5)}$$

Then (B4) takes the form

$$\sum_j Y_{-\alpha,-j}(\lambda_{-\beta})A_{-j,-i}(\lambda_{-\beta}) = \lambda_\alpha Y_{-\alpha,-i}(\lambda_{-\beta}).$$

This result holds for all α, i and j is a dummy index. Thus we can change the sign of all of them to obtain

$$\sum_j Y_{\alpha j}(\lambda_{-\beta})A_{ji}(\lambda_{-\beta}) = \lambda_{-\alpha} Y_{\alpha i}(\lambda_{-\beta}). \quad \text{(B6)}$$

This shows that Y's are the left eigenvectors. Thus we can obtain a left eigenvector from a right eigenvector. To obtain the left eigenvector corresponding to the eigenvalue α, we must take the right eigenvector corresponding to the eigenvalue $(-\alpha)$ and reverse its components and replace every λ_β appearing in it by $\lambda_{-\beta}$. Since reversing the components does not alter the sum,

1615   J. Math. Phys., Vol. 17, No. 8, August 1976

C.O. Nwachuku and M.A. Rashid   1615

to compute the sum of the components of a left eigenvector corresponding to the eigenvalue λ_α, we just have to replace all λ_β by $\lambda_{-\beta}$ in the sum for the right eigenvector corresponding to the eigenvalue $\lambda_{-\alpha}$.

*On leave of absence from Department of Mathematics, University of Benin, Nigeria.

†On leave of absence from Department of Mathematics, University of Islamabad, Pakistan.

[1] A. M. Perelomov and V. S. Popov, Sov. J. Nucl. Phys. 3, 676 (1966).

[2] A. M. Perelomov and V. S. Popov, Sov. J. Nucl. Phys. 3, 819 (1966).

[3] V. S. Popov and A. M. Perelomov, Sov. J. Nucl. Phys. 5, 489 (1967).

[4] J. D. Louck, Am. J. Phys. 38, 3 (1970).

[5] J. D. Louck and L. C. Biedenharn, J. Math. Phys. 11, 2368 (1970).

[6] S. Okubo, J. Math. Phys. 16, 528 (1975).

[7] M. K. F. Wong and H. Y. Yeh, J. Math. Phys. 16, 1239 (1975).

[8] Our list of references is not exhaustive. A better list appears in Ref. 7 and the Introduction in this reference contains information on the contents of these references.

[9] See, for example, Eq. (6) in Ref. 3.

[10] See Eq. (4.13) in Ref. 7.

[11] See Eq. (3.1) of the present paper.

[12] Note that the table in Ref. 2 for this matrix uses the usual ordering. Our table will be obtained by replacing $f_1, \ldots, f_n$ by $f_n, f_{n-1}, \ldots, f_1$ and changing the indexing from $1, 2, \ldots, n, 0, -(n-1), \ldots, -1$ to $n, n-1, \ldots, 0, -1, \ldots, -n$. Our ordering follows the usual ordering of decreasing indices.

[13] See Eq. (6) in Ref. 3. This result also appears in Refs. 6 and 7.

ERRATA

Erratum: On the equality of S operators corresponding to unitarily equivalent Hamiltonians in single channel scattering [J. Math. Phys. 17, 409 (1976)]

A. W. Sáenz and W. W. Zachary

Naval Research Laboratory, Washington D. C.
(Received 27 April 1976)

The first sentence of the first paragraph immediately following Eqs. (5.12) should read:

Invoking (5.7b), (5.7d), (5.11), (5.12), and the compactness of Γ_1^c, we conclude that $\sigma_a(\mathbf{y}, t) \to 1$ and $\lambda_a(\mathbf{y}, t)/(1 + 4t^2) \to 0$ for given $\mathbf{0} \neq \mathbf{y} \in R^3$, $\mathbf{a} \in R^3$ as $t \to \pm \infty$.

CUMULATIVE AUTHOR INDEX

to compute the sum of the components of a left eigenvector corresponding to the eigenvalue λ_α, we just have to replace all λ_β by $\lambda_{-\beta}$ in the sum for the right eigenvector corresponding to the eigenvalue $\lambda_{-\alpha}$.

*On leave of absence from Department of Mathematics, University of Benin, Nigeria.

†On leave of absence from Department of Mathematics, University of Islamabad, Pakistan.

[1] A. M. Perelomov and V.S. Popov, Sov. J. Nucl. Phys. 3, 676 (1966).

[2] A. M. Perelomov and V.S. Popov, Sov. J. Nucl. Phys. 3, 819 (1966).

[3] V. S. Popov and A.M. Perelomov, Sov. J. Nucl. Phys. 5, 489 (1967).

[4] J. D. Louck, Am. J. Phys. 38, 3 (1970).

[5] J. D. Louck and L.C. Biedenharn, J. Math. Phys. 11, 2368 (1970).

[6] S. Okubo, J. Math. Phys. 16, 528 (1975).

[7] M. K. F. Wong and H.Y. Yeh, J. Math. Phys. 16, 1239 (1975).

[8] Our list of references is not exhaustive. A better list appears in Ref. 7 and the Introduction in this reference contains information on the contents of these references.

[9] See, for example, Eq. (6) in Ref. 3.

[10] See Eq. (4.13) in Ref. 7.

[11] See Eq. (3.1) of the present paper.

[12] Note that the table in Ref. 2 for this matrix uses the usual ordering. Our table will be obtained by replacing $f_1, \ldots, f_n$ by $f_n, f_{n-1}, \ldots, f_1$ and changing the indexing from $1, 2, \ldots, n, 0, -(n-1), \ldots, -1$ to $n, n-1, \ldots, 0, -1, \ldots, -n$. Our ordering follows the usual ordering of decreasing indices.

[13] See Eq. (6) in Ref. 3. This result also appears in Refs. 6 and 7.

ERRATA

Erratum: On the equality of S operators corresponding to unitarily equivalent Hamiltonians in single channel scattering [J. Math. Phys. 17, 409 (1976)]

A. W. Sáenz and W. W. Zachary

Naval Research Laboratory, Washington D. C.
(Received 27 April 1976)

The first sentence of the first paragraph immediately following Eqs. (5.12) should read:

Invoking (5.7b), (5.7d), (5.11), (5.12), and the compactness of Γ_1^c, we conclude that $\sigma_a(\mathbf{y}, t) \to 1$ and $\lambda_a(\mathbf{y}, t)/(1 + 4t^2) \to 0$ for given $0 \neq \mathbf{y} \in R^3$, $\mathbf{a} \in R^3$ as $t \to \pm\infty$.

CUMULATIVE AUTHOR INDEX