Gauge invariant canonical mechanics for charged particles

R. J. Torrence and W. M. Tulczyjew

The University of Calgary, Calgary, Alberta, Canada
(Received 1 March 1973)

The mechanics of charged particle motion is presented in a five-dimensional form compatible with the five-dimensional Kaluza theory of the electromagnetic field. Gauge dependence is given an intrinsic geometrical interpretation and the role of generalized momenta is clarified. The theory is a new example of a mechanical system with constraints and offers an interesting exercise in canonical quantization with a nonstandard Poisson bracket.

1. INTRODUCTION

The dynamics of a charged particle in standard canonical form requires the use of a gauge-dependent momentum. The gauge dependence of the electromagnetic potential has been clarified by giving it a geometrical interpretation in the five-dimensional Kaluza theory of the electromagnetic field,[1-3] and the purpose of this paper is to show that a similar clarification of the gauge dependence of momentum is possible. This clarification suggests a modification of canonical dynamics in which a different, gauge-independent, momentum is used, at the expense of a modification of the Poisson bracket. It is shown that in all respects, including canonical quantization, the modified Poisson bracket can replace the standard version.

The geometry of the five-dimensional theory of Kaluza is reviewed in Sec. 2. In comparison with the formulations found in standard texts, more emphasis is given to the intrinsic features of the five-dimensional geometry and to the reduction of five-dimensional quantities to four dimensions. The reduction of some quantities results in classes of objects related by gauge transformations, and the geometric origin of gauge transformations is made clear.

Section 3 contains the formulation of the dynamics of charged particles in five dimensions and its reduction to four dimensions. The reduction leads in a natural way to a gauge-independent canonical theory with a nonstandard Poisson bracket. The geometric meaning of the usual gauge-dependent momenta is discussed. The five-dimensional dynamics presented is an example of a dynamics with primary constraints and no secondary constraints.[4-6]

In the last section the canonical quantization of the five- and four-dimensional dynamical systems is given. In the case of the five-dimensional theory we have an example of the quantization of a mechanical system with constraints. The absence of secondary constraints in a classical theory does not imply that there are none in the corresponding quantum theory, and in fact they do arise in this case. Since we feel that the problem of secondary constraints of a quantum-mechanical system is far from solved we prefer to avoid them here. Fortunately the assumption that the Einstein and Maxwell field equations hold in vacuum eliminates any secondary constraints. The manner of appearance of secondary constraints in this example, and their relation to the field equations, is in our opinion of sufficient interest to justify a separate investigation. In the canonical quantization of the four-dimensional theory the usual procedure must be modified because of the nonstandard Poisson bracket.

2. KALUZA THEORY

A. The mathematical structure

We assume a Riemannian manifold Z of five dimensions and signature $(+,+,-,-,-)$, a Riemannian manifold X of four dimensions and signature $(+,-,-,-)$, and the one-dimensional Riemannian manifold of real numbers Y. Coordinates of Z and X are represented by z^α and x^a, respectively, while y denotes an element of Y.[7] We assume a mapping Π from Z onto X which, in terms of coordinates, we represent by the functions[8]

$$x^a = \Pi^a(z^\alpha). \tag{2.1}$$

We assume a *class* Φ of mappings, $\varphi, \varphi', \ldots$, from Z onto Y which, in terms of coordinates, we represent by the class of functions[8]

$$\begin{aligned} y &= \varphi(z^\alpha), \\ y &= \varphi'(z^\alpha), \\ &\vdots \\ &\vdots \end{aligned} \tag{2.2}$$

It is clear that Π and any one φ together provide a mapping from Z to $X \times Y$ by assigning to each z^α the element $(\Pi^a(z^\alpha), \varphi(z^\alpha))$ of $X \times Y$. We assume that for each φ in Φ the mapping from Z to $X \times Y$ so constructed is one-to-one and onto. The significance of this assumption for the mappings Π and φ is pictured in Fig. 1 (where Z and X each have three dimensions suppressed). Sets of elements of Z mapped into a single

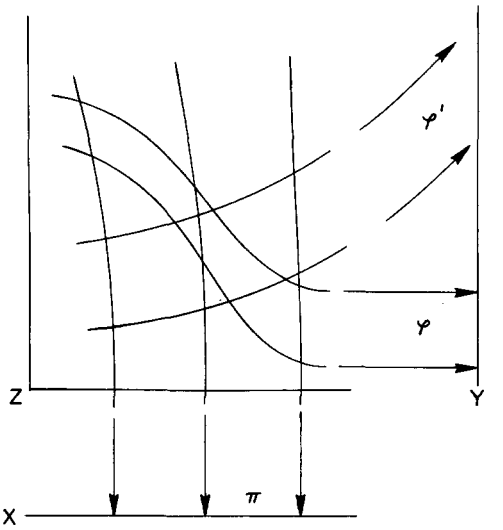


FIG. 1. The manifolds Z, X, and Y, the mapping Π, and two sectionings φ and φ', of Φ.

element of X by Π are referred to as *fibers* of Z, and are indicated by the vertical line segments in Fig. 1. Since there is only one mapping Π, the manifold Z has only one set of fibers. Sets of elements of Z mapped to a single element of Y by a mapping φ of Φ are referred to as *sections* of Z, with the set of sections corresponding to a particular φ of Φ called a *sectioning* of Z. Sections of Z are indicated by the curves transverse to fibers in Fig. 1. Since Φ has many members, Z has many sectionings, and in Fig. 1 two possible sectionings are indicated. The nature of the mappings given by Eqs. (2.1) and (2.2) insure that fibers do not intersect themselves or one another, and that sections do not intersect themselves or one another; that the mapping Z to $X \times Y$ described above is one-to-one and onto implies that each fiber intersects each section at exactly one point. Finally, we assume a one parameter group of transformations which act on Z and which preserve the fibers and sectionings, that is if we represent these transformations by

$$z^{\alpha'} = \gamma^{\alpha}(z^{\beta}, \gamma),$$

where γ is a real parameter chosen so that

$$\gamma^{\alpha}(\gamma^{\beta}(z^{\alpha}, \gamma), \gamma') = \gamma^{\alpha}(z^{\alpha}, \gamma + \gamma'),$$

then we assume that

$$\Pi^{\alpha}(\gamma^{\alpha}(z^{\beta}, \gamma)) = \Pi^{\alpha}(z^{\beta}) \tag{2.3}$$

and

$$\varphi(\gamma^{\alpha}(z^{\beta}, \gamma)) = \varphi(z^{\beta}) + \gamma. \tag{2.4}$$

If the one-to-one and onto property of the mapping from Z to $X \times Y$ is taken into account these equations imply that any fiber can be generated by the one parameter group acting on any element of that fiber,[9] and all sections of a particular sectioning of Z can be generated by the one parameter group acting on any one section in that sectioning. In fact, it is further implied that having one sectioning, φ, of Z one can obtain the entire class, Φ, by adding to φ arbitrary functions on Z constant along fibers, since Eq. (2.4) implies that the difference between any two sectionings is constant along fibers. The action of the one parameter group is pictured in Fig. 2. Equations (2.3) and (2.4) have differential forms. If we define

$$A^{\alpha} = \frac{\partial \gamma^{\alpha}(z^{\beta}, \gamma)}{\partial \gamma}\bigg|_{\gamma=0},$$

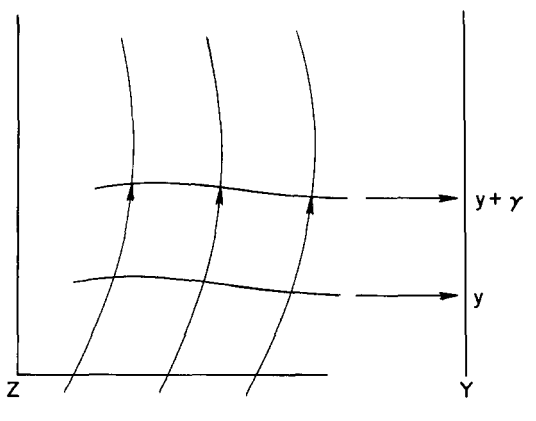


FIG. 2. The action of the element γ of the one parameter group on fibers and sections of Z.

the differentiation of Eqs. (2.3) and (2.4) with respect to γ yields

$$A^{\alpha} \frac{\partial \Pi^{a}}{\partial z^{\alpha}} = 0 \tag{2.5}$$

and

$$A^{\alpha} \frac{\partial \varphi}{\partial z^{\alpha}} = 1, \tag{2.6}$$

respectively. Clearly A^{α} is the field of vectors tangent to the fibers of Z.

There remain conditions on the metric structures of Z, X, and Y to be stated. We assume that the metric tensor of Z, $g_{\alpha\beta}$, satisfies[10]

$$\underset{A}{\pounds} g_{\alpha\beta} = 0, \tag{2.7}$$

which means that the one parameter group acts on Z as a one parameter group of motions, the fibers of Z being the streamlines of the one parameter group. The last assumption in specifying our mathematical structure is

$$g_{\alpha\beta}(z\gamma) = A_{\alpha} A_{\beta} + g_{ab}(\Pi^{c}(z\gamma)) \frac{\partial \Pi^{a}}{\partial z^{\alpha}} \frac{\partial \Pi^{b}}{\partial z^{\beta}}, \tag{2.8}$$

where $g_{ab}(x^c)$ is the metric tensor of X. This assumption decomposes, in a sense to be made precise, the metric of Z into the metric of Y and the metric of X.

A particular orthogonal decomposition of tensors on Z plays an important role in what follows. The decomposition consists of projections, at each point, onto and orthogonal to the fiber direction at that point. If the set of fibers is hypersurface orthogonal, i.e., if $A_{\alpha,\beta} - A_{\beta,\alpha} = 0$, then the projections orthogonal to the fiber direction can be defined as onto four-dimensional hypersurfaces everywhere orthogonal to the fibers. However, if we have $A_{\alpha,\beta} - A_{\beta,\alpha} \neq 0$ then no such hypersurfaces exist.

The mathematics of projection operators is generally known, but a few results are given here. If we define

$$\epsilon^{\alpha}_{\beta} = \delta^{\alpha}_{\beta} - A^{\alpha} A_{\beta}, \tag{2.9}$$

then $V^{\alpha}_{\parallel} \equiv A^{\alpha} A_{\beta} V^{\beta}$ and $V^{\alpha}_{\perp} \equiv \epsilon^{\alpha}_{\beta} V^{\beta}$ are the projections of a vector, V^{α}, onto and orthogonal to the fiber direction. There is an alternate way of defining essentially the same thing. If we define A^{α}_{a} uniquely by

$$\frac{\partial \Pi^{b}}{\partial z^{\alpha}} A^{\alpha}_{a} = \delta^{b}_{a}, \quad A_{\alpha} A^{\alpha}_{a} = 0, \tag{2.10}$$

then $\nu \equiv A^{\alpha} V_{\alpha}$ and $\nu_{a} = A^{\alpha}_{a} V_{\alpha}$ [$\nu^{a} = (\partial \Pi^{a}/\partial z^{\alpha}) A^{\alpha}$] can also be called the projections onto and orthogonal to the fiber direction of V_{α}. (The generalization of these definitions to tensors is obvious.) If we define

$$A^{\alpha}_{h} = (A^{\alpha}_{a}, A^{\alpha}), \quad A^{h}_{\alpha} = \left(\frac{\partial \Pi^{a}}{\partial z^{\alpha}}, A_{\alpha}\right), \tag{2.11}$$

the quantities (ν^a, ν) just defined can be interpreted as the components of V^{α} with respect to an anholonomic reference frame[11] related to the (holonomic) reference frame z^{α} by A^{α}_{h} and A^{h}_{α}. The connections between $V^{\alpha}_{\parallel}$ and $V^{\alpha}_{\perp}$, and ν and ν^a, are $V^{\alpha}_{\parallel} = \nu A^{\alpha}$ and $V^{\alpha}_{\perp} = \nu^{a} A^{\alpha}_{a}$ [$\nu^{a} = (\partial \Pi^{a}/\partial z^{\alpha}) V^{\alpha}_{\parallel}$]. The essential difference is that V^{α} is a vector on Z while ν^a is not, although it is defined on Z. If it can be shown that $\underset{A}{\pounds} \nu^{a} = 0$ it follows from the theorem in the appendix that the ν^a are constant along fibers and thus that

$$\nu^a(z^\alpha) = \nu^a(\Pi^b(z^\alpha)) = \nu^a(x^b).$$

If we do a coordinate transformation $x^{a'} = x^{a'}(x^b)$ on X then $x^{a'}(x^b) = x^{a'}(x^b(z^\alpha)) \equiv \Pi^{a'}(z^\alpha)$ represents the mapping Π, so $\nu^{b'} = (\partial\Pi^{b'}/\partial z^\alpha)V^\alpha$, which means that $\nu^{b'} = (\partial x^{b'}/\partial x^a)(\partial\Pi^a/\partial z^\alpha)\ V^\alpha = (\partial x^{b'}/\partial x^a)\nu^a$. Thus the ν^a are defined on X as well as Z, and transform as the components of a vector with respect to coordinate transformations on X.

The assumption about $g_{\alpha\beta}$ embodied in Eq. (2.8) has the significance there stated in the sense that $g_{\alpha\beta}$ decomposes orthogonally into $g_{\alpha\beta}A^\alpha A^\beta = 1$, $g_{\alpha\beta}A^\alpha A^\beta_a = 0$, and $g_{\alpha\beta}A^\alpha_a A^\beta_b = g_{ab}$.

B. The physical interpretation

The five-dimensional Riemannian manifold has defined on it the vector field A_α and from A_α we form

$$F_{\alpha\beta} = A_{\alpha,\beta} - A_{\beta,\alpha}. \tag{2.12}$$

From this tensor defined on Z we can obtain a tensor f_{ab} defined on X; the general method was outlined in the preceding section. We first define

$$f_{ab}(z^\alpha) = A^\alpha_a A^\beta_b F_{\alpha\beta}.$$

We have $\underset{A}{\pounds} F_{\alpha\beta} = (\underset{A}{\pounds} A_\alpha)_{,\beta} - (\underset{A}{\pounds} A_\beta)_{,\alpha}$ since Lie and partial differentiation commute, and since $\underset{A}{\pounds} A_\alpha = 0$ for any A_α, we have $\underset{A}{\pounds} F_{\alpha\beta} = 0$. Since[12] $\underset{A}{\pounds} A^\alpha_a = 0$ also it follows that $\underset{A}{\pounds} f_{ab} = 0$. By the theorem in the appendix this means that $f_{ab,\gamma}A^\gamma = 0$, that is f_{ab} is constant along fibers, and we can conclude that $f_{ab}(z^\alpha) = f_{ab}(\Pi^c(z^\alpha)) = f_{ab}(x^c)$ and that f_{ab} is in fact a tensor defined on X.

The physical interpretation of the mathematical structure follows from identifying X and g_{ab} with physical space-time and its metric structure, and f_{ab} with the eletromagnetic field tensor.

C. Gauge transformations

The vector field A^α plays a double role in the above structure. On one hand it is the field of vectors tangent to the streamlines of the one parameter group, and as such is unique. On the other hand, the covariant form A_α serves as a vector potential for the electromagnetic field [Eq. (2.12)], and in this role it is not unique; if we define $\bar{A}_\alpha = A_\alpha - \chi_{,\alpha}$, where χ is an arbitrary scalar field on Z, then $F_{\alpha\beta} = A_{\alpha,\beta} - A_{\beta,\alpha} = \bar{A}_{\alpha,\beta} - \bar{A}_{\beta,\alpha}$. In a sense there exists a class of five-dimensional gauge transformations, however there is a "preferred gauge" defined by $\bar{A}_\alpha = A_\alpha$. It is natural to expect that as A_α serves as a vector potential for $f_{\alpha\beta}$, and f_{ab} is the reduction of $f_{\alpha\beta}$ to X, it should be possible to reduce A_α to a vector field, a_a, on X which will serve as the familiar four vector potential for f_{ab}. However, if we start by defining $a_a \equiv A^\alpha_a A_\alpha$ it follows from Eq. (2.10) that $a_a = 0$, and the construction breaks down. A successful construction can be based on the following argument. If a five-dimensional quantity is to be essentially four-dimensional its projection onto the fiber direction with respect to any index should vanish. Thus, if we check $f_{\alpha\beta}$ we have $F_{\alpha\beta}A^\beta = A_{\alpha;\beta}A^\beta - A_{\beta;\alpha}A^\beta = 0$, since $A_\alpha A^\alpha = 1$ implies that

$$A_{\alpha;\beta}A^\alpha = 0, \tag{2.13}$$

and $g_{\alpha\beta} = A_{\alpha;\beta} + A_{\beta;\alpha} = 0$ combined with Eq. (2.13) implies that

$$A_{\beta;\alpha}A^\alpha = 0. \tag{2.14}$$

By the same criterion $A_\alpha A^\alpha = 1$ suggests that A_α itself is not a good candidate for "reduction" to X. If however, we consider $\bar{A}_\alpha = A_\alpha - \chi_{,\alpha}$, and choose χ to be any $\varphi(z^\alpha)$ of the class Φ, then $A^\alpha\varphi_{,\alpha} = 1$ and $\bar{A}_\alpha A^\alpha = 0$. Thus the set of five-dimensional vector potentials $\Phi_\alpha \equiv A_\alpha - \varphi_{,\alpha}$ are candidates for reduction. Defining

$$\phi_a = A^\alpha_a \Phi_\alpha, \tag{2.15}$$

we have

$$\underset{A}{\pounds}\phi_a = -\underset{A}{\pounds}(A^\alpha_a\varphi_{,\alpha}) = -A^\alpha_a \underset{A}{\pounds}\varphi_{,\alpha} = 0,$$

since

$$\underset{A}{\pounds}\varphi_{,\alpha} = \varphi_{,\alpha\beta}A^\beta + \varphi_{,\beta}A^\beta{}_{,\alpha} = \varphi_{,\beta\alpha}A^\beta + \varphi_{,\beta}A^\beta{}_{,\alpha} = 0,$$

as $\varphi_{,\beta}A^\beta = 1$. Thus by the theorem in the appendix

$$\phi_a(z^\alpha) = \phi_a(x^c)$$

is a vector field on X. It is easy to verify that ϕ_a does serve as a potential for f_{ab}; we have $f_{ab} = A^\alpha_a A^\beta_b F_{\alpha\beta}$, where $F_{\alpha\beta} = \Phi_{\alpha,\beta} - \Phi_{\beta,\alpha}$ and $\Phi_\alpha = (\partial\Pi^a/\partial z^\alpha)\phi_a$, and this leads directly to $f_{ab} = \phi_{a,b} - \phi_{b,a}$. If instead of φ we use any other sectioning $\varphi' = \varphi + \psi$, where $\psi(z^\alpha)$ is constant along fibers, then $\varphi'_{,\alpha}A^\alpha = 1$ is preserved, and the same construction yields $f_{ab} = \phi'_{a,b} - \phi'_{b,a}$, where $\phi'_a = \phi_a + \psi_{,a}$. Consequently, the freedom of using any sectioning, φ, of the class Φ is precisely the usual four-dimensional gauge freedom.

3. CLASSICAL DYNAMICS OF CHARGED PARTICLES

A. Five-dimensional formulation

We formulate a five-dimensional canonical mechanics of charged particles of charge e and mass m which contains the correct four-dimensional mechanics. We start with the action

$$A = \int\{m[(g_{\alpha\beta} - A_\alpha A_\beta)dz^\alpha dz^\beta]^{1/2} + eA_\alpha dz^\alpha\}.$$

If the trajectory is parametrized in terms of an arbitrary parameter this can be rewritten as

$$A = \int\left[m\left((g_{\alpha\beta} - A_\alpha A_\beta)\frac{dz^\alpha}{d\sigma}\frac{dz^\beta}{d\sigma}\right)^{1/2} + eA_\alpha\frac{dz^\alpha}{d\sigma}\right]d\sigma,$$

with the corresponding Lagrangian

$$L = m\left((g_{\alpha\beta} - A_\alpha A_\beta)\frac{dz^\alpha}{d\sigma}\frac{dz^\beta}{d\sigma}\right)^{1/2} + eA_\alpha\frac{dz^\alpha}{d\sigma}.$$

If we define

$$P_\alpha \equiv \frac{\partial L}{\partial(dz^\alpha/d\sigma)} = \left((g_{\gamma\delta} - A_\gamma A_\delta)\frac{dz^\gamma}{d\sigma}\frac{dz^\delta}{d\sigma}\right)^{-1/2} \times m(g_{\alpha\beta} - A_\alpha B_\beta)\frac{dz^\beta}{d\sigma} + eA_\gamma$$

it is obvious that

$$\Phi_1 \equiv P_\alpha A^\alpha - e = 0, \tag{3.1}$$

and easy to show, using Eq. (3.1), that

$$\Phi_2 \equiv [(g^{\alpha\beta} - A^\alpha A^\beta)P_\alpha P_\beta]^{1/2} - m = [g^{\alpha\beta} - (P_\alpha - eA_\alpha)(P_\beta - eA_\beta)]^{1/2} - m = 0. \tag{3.2}$$

Thus the theory has constraints. It follows from $\Phi_1 = 0$ that the component of P_α in the direction of the fiber is the charge. Using Eq. (2.8) we can rewrite $\Phi_2 = 0$ as $(g^{ab}A^\alpha_a P_\alpha A^\beta_b P_\beta)^{1/2} = m$, and this implies that the projection of P_α orthogonal to the fiber direction is the energy–momentum 4-vector. Equations (3.3) and (3.4) are primary[4-6] constraints of the Hamiltonian formalism we are constructing. If follows that the Hamiltonian is not unique, and is given by $H = P_\alpha(dz^\alpha/d\sigma) - L + c_1\Phi_1 + c_2\Phi_2$, where c_1 and c_2 are arbitrary functions of z^α and P_α. Substituting for P_α from its definition we have $P_\alpha(dz^\alpha/d\sigma) - L = 0$, consequently $H = c_1\Phi_1 + c_2\Phi_2$ is the full Hamiltonian. It vanishes modulo the constraints as it should. Checking for secondary constraints we see that $[\Phi_1, H] = [\Phi_2, H] = 0$ if $[\Phi_1, \Phi_2] = 0$, where the bracket is the Poisson bracket. We have

$$\begin{aligned}[\Phi_1, \Phi_2] &= [A^\alpha P_\alpha, (g^{\beta\gamma} - A^\beta A^\gamma)P_\beta P_\gamma] \\ &= 2A^\alpha{}_{,\gamma} P_\alpha (g^{\beta\gamma} - A^\beta A^\gamma)P_\beta \\ &\quad - A^\gamma(g^{\alpha\beta}{}_{,\gamma} - A^\alpha{}_{,\gamma}A^\beta - A^\alpha A^\beta{}_{,\gamma})P_\alpha P_\beta,\end{aligned}$$

where $\Phi_2 = 0$ has been used. Since

$$\underset{A}{\pounds} g^{\alpha\beta} = 0 = g^{\alpha\beta}{}_{,\gamma}A^\gamma - g^{\gamma\beta}A^\gamma{}_{,\gamma} - g^{\alpha\gamma}A^\beta{}_{,\gamma},$$

it follows that

$$\begin{aligned}[\Phi_1, \Phi_2] = P_\alpha P_\beta [&2A^\alpha{}_{,\gamma}(g^{\beta\gamma} - A^\beta A^\gamma) \\ &- (g^{\alpha\gamma}A^\beta{}_{,\gamma} + g^{\gamma\beta}A^\alpha{}_{,\gamma} - A^\gamma A^\alpha{}_{,\gamma}A^\beta - A^\gamma A^\alpha A^\beta{}_{,\gamma})].\end{aligned}$$

The square bracket is antisymmetric in α and β, therefore we have $[\Phi_1, \Phi_2] = 0$ and there are no secondary constraints. As a consequence we can write the Hamiltonian as

$$H = \nu_1\Phi_1 + \nu_2\Phi_2, \tag{3.3}$$

where ν_1 and ν_2 are arbitrary functions of σ, and the canonical equations of motion are

$$\begin{aligned}\frac{dP_\alpha}{d\sigma} &= [P_\alpha, H] = -\nu_1(\sigma)\frac{\partial\Phi_1}{\partial z^\alpha} - \nu_2(\sigma)\frac{\partial\Phi_2}{\partial z^\alpha}, \\ \frac{dz^\alpha}{d\sigma} &= [z^\alpha, H] = \nu_1(\sigma)\frac{\partial\Phi_1}{\partial P_\alpha} + \nu_2(\sigma)\frac{\partial\Phi_2}{\partial P_\alpha}.\end{aligned} \tag{3.4}$$

Equations (3.4) can be put into a more convenient form. They are clearly satisfied by $z^\alpha(\tau, \lambda)$ and $P_\alpha(\tau, \lambda)$, where $\tau = \int\nu_2(\sigma)d\sigma$ and $\lambda = \int\nu_1(\sigma)d\sigma$, if

$$\begin{aligned}\frac{\partial P_\alpha}{\partial\lambda} &= [P_\alpha, \Phi_1], \quad \frac{\partial P_\alpha}{\partial\tau} = [P_\alpha, \Phi_2], \\ \frac{\partial z^\alpha}{\partial\lambda} &= [z^\alpha, \Phi_1], \quad \frac{\partial z^\alpha}{\partial\tau} = [z^\alpha, \Phi_2].\end{aligned} \tag{3.5}$$

The consistency of Eqs. (3.5) follows from the equality of the mixed second partial derivatives of z^α and P_α, which is easily established. We can then take Eqs. (3.5), which contain no arbitrary functions, as our equations of motion in place of Eqs. (3.4). It follows that a "trajectory" in Z is a two-dimensional surface in Z parametrized by τ and λ.

B. Reduction to four dimensions

We wish now to consider the four-dimensional implications of the five-dimensional formalism just presented. The importance of the Poisson bracket,

$$[A, B] \equiv \frac{\partial A}{\partial z^\alpha}\frac{\partial B}{\partial P_\alpha} - \frac{\partial B}{\partial z^\alpha}\frac{\partial A}{\partial P_\alpha}, \tag{3.6}$$

where A and B are arbitrary (differentiable) functions of z^α and P_α, suggests that we start by considering *its* reduction. The functions A and B should be thought of as being defined on a ten-dimensional phase space coordinatized by z^α and P_α. The four-dimensional momentum is of course

$$p_a \equiv A^\alpha_a P_\alpha, \tag{3.7}$$

and we now consider an eight-dimensional phase space coordinatized by x^a and p_a. We specialize the form of the Poisson bracket defined by Eq. (3.6) to the case when A and B depend on z^α and P_α only through $x^a \equiv \Pi^a(z^\alpha)$ and $p_a = A^\alpha_a P_\alpha$, that is when A and B are defined on the eight-dimensional phase space mentioned above. Since we are assuming that $A = A(x^a, p_a) = A(\Pi^a(z^\alpha), A^\alpha_a P_\alpha)$, we have

$$\begin{aligned}\frac{\partial A}{\partial z^\alpha} &= \frac{\partial A}{\partial x^a}\frac{\partial\Pi^a}{\partial z^\alpha} + \frac{\partial A}{\partial p_a}A^\beta_{a,\alpha}P_\beta, \\ \frac{\partial A}{\partial P_\alpha} &= \frac{\partial A}{\partial p_a}A^\alpha_a;\end{aligned} \tag{3.8}$$

and similarly for B. Rewriting the Poisson bracket as

$$\frac{\partial A}{\partial z^\alpha}\delta^\alpha_\beta\frac{\partial B}{\partial P_\beta} - \frac{\partial B}{\partial z^\alpha}\delta^\alpha_\beta\frac{\partial A}{\partial P_\alpha}$$

and substituting

$$\delta^\alpha_\beta = A^\alpha_a\frac{\partial\Pi^a}{\partial z^\beta} + A^\alpha A_\beta$$

and Eqs. (3.8), we obtain

$$\begin{aligned}[A, B] = &\left(\frac{\partial A}{\partial x^a}\frac{\partial\Pi^a}{\partial z^\alpha} + \frac{\partial A}{\partial p_a}A^\beta_{a,\alpha}P_\beta\right)\left(A^\alpha_b\frac{\partial\Pi^b}{\partial z^\gamma} + A^\alpha A_\gamma\right)\frac{\partial B}{\partial p_c}A^\gamma{}_c \\ &-\left(\frac{\partial B}{\partial x^a}\frac{\partial\Pi^a}{\partial z^\alpha} + \frac{\partial B}{\partial p_a}A^\beta_{a,\alpha}P_\beta\right)\left(A^\alpha_b\frac{\partial\Pi^b}{\partial z^\gamma} + A^\alpha A_\gamma\right)\frac{\partial A}{\partial p_c}A^\gamma{}_c.\end{aligned}$$

Expanding the products and using the defining equations of the A^α_a yields

$$\begin{aligned}[A, B] = &\frac{\partial A}{\partial x^a}\frac{\partial B}{\partial p_a} + \frac{\partial A}{\partial p_a}A^\beta_{a,\alpha}P_\beta A^\alpha_b\frac{\partial B}{\partial p_b} \\ &- \frac{\partial B}{\partial x^a}\frac{\partial A}{\partial p_a} - \frac{\partial B}{\partial p_a}A^\beta_{a,}P_\beta A^\alpha_b\frac{\partial A}{\partial p_b}.\end{aligned}$$

It follows from $p_a = A^\alpha_a P_\alpha$ that $P_\alpha = (\partial\Pi^a/\partial z^\alpha)p_a + eA_\alpha$; substituting this into the last result we easily obtain

$$[A, B] = \frac{\partial A}{\partial x^a}\frac{\partial B}{\partial p_a} - \frac{\partial B}{\partial x^a}\frac{\partial A}{\partial p_a} + e\frac{\partial B}{\partial p_a}f_{ab}\frac{\partial A}{\partial p_b}, \tag{3.9}$$

where Eqs. (2.10) have been used again.

Let us now consider a charged particle with a trajectory in Z given by $z^\alpha(\tau, \lambda)$, and a 5-momentum given by $P_\alpha(\tau, \lambda)$, where z^α and P_α satisfy Eqs. (3.5). The space-time trajectory (the trajectory reduced to X) is $x^a(\tau, \lambda) = \Pi^a(z^\alpha(\tau, \lambda))$, while $p_a = A^\alpha_a P_\alpha$ is the natural 4-momentum. We have

$$\frac{\partial x^a}{\partial\lambda} = \frac{\partial\Pi^a}{\partial z^\alpha}\frac{\partial z^\alpha}{\partial\lambda} = \frac{\partial\Pi^a}{\partial z^\alpha}[z^\alpha, \Phi_1] = \frac{\partial\Pi^a}{\partial z^\alpha}A^\alpha = 0,$$

and

$$\frac{\partial p_a}{\partial \lambda} = A_a^\alpha[P_\alpha, \Phi_1] + [A_a^\alpha, \Phi_1]P_\alpha$$
$$= A^\beta{}_{,\alpha} P_\beta A_a^\alpha + P_\alpha A_{a,\beta}^\alpha A^\beta$$
$$= P_\alpha \pounds_A A_a^\alpha = 0.$$

Thus we have $x^a(\tau, \lambda) = x^a(\tau), p_a(\tau, \lambda) = p_a(\tau)$. From Eqs. (3.5) it follows that $\partial x^a/\partial\tau = [x^a, \Phi_2]$ and $\partial p_a/\partial\tau = [p_a, \Phi_2]$.

It follows from Eq. (2.8) that $\Phi_2 = [p_a p_b g^{ab}]^{1/2} - m$, thus the above brackets are really between functions defined on the eight-dimensional phase space. Consequently, we can apply Eq. (3.9) and immediately obtain

$$\frac{\partial x^a}{\partial \tau} = \frac{\partial \Phi_2}{\partial p_a}, \tag{3.10a}$$

$$\frac{\partial p_a}{\partial \tau} = -\frac{\partial \Phi_2}{\partial x^a} + e f_{ba}\frac{\partial \Phi_2}{\partial p_b}. \tag{3.10b}$$

Substituting $[g^{ab}p_a p_b]^{1/2} - m$ for Φ_2 in these equations we obtain

$$\frac{\partial x^a}{\partial \tau} = \frac{g^{ab}p_b}{m}, \tag{3.11a}$$

$$\frac{\partial p_a}{\partial \tau} = -\frac{1}{2m} g^{bc}{}_{,a} p_b p_c + \frac{e}{m} f_{ba} g^{bc} p_c. \tag{3.11b}$$

Eliminating the p_a from Eq. (3.11b) by means of Eqs. (3.11a) gives

$$\frac{d^2x^b}{d\tau^2} = -\left\{ {b \atop cd} \right\} \frac{dx^c}{d\tau}\frac{dx^d}{d\tau} + \frac{e}{m} g^{bc} f_{ac} \frac{dx^c}{d\tau}, \tag{3.12}$$

where $\left\{ {b \atop cd} \right\}$ is the Christoffel symbol of the four-dimensional Riemannian manifold X. Since Eq. (3.12) is precisely the equation of motion of a charged particle in a combined gravitational and electromagnetic field, we have verified that the five-dimensional dynamics postulated does contain the correct four-dimensional dynamics.

Going back to Eq. (3.11a) we see that p_a, the natural 4-momentum introduced by Eq. (3.7), is the *ordinary* momentum of the charged particle rather than the generalized momentum; thus Eqs. (3.11) are the equations of motion of the charged particle in terms of its ordinary momentum.

It is not difficult to show directly that if we eliminate p_a from Eqs. (3.10) in favor of

$$m_a \equiv p_a + e\phi_a,$$

where $\phi_a = -A_a^\alpha \varphi_{,\alpha}$ is the four-dimensional vector potential defined by Eq. (2.15), we obtain

$$\frac{\partial x^a}{\partial \tau} = \frac{\partial \Phi_2}{\partial m_a}, \qquad \frac{\partial m_a}{\partial \tau} = -\frac{\partial \Phi_2}{\partial x^a}. \tag{3.13}$$

Here Φ_2 should be expressed in terms of m_a, i.e.,

$$\Phi_2 = [g^{ab}(m_a - e\phi_a)(m_b - e\phi_b)]^{1/2} - m.$$

This was of course to be expected. The m_a is the usual generalized momentum of a charged particle and Eqs. (3.13) are the usual canonical equations of motion which can be written using the usual Poisson bracket. The transformation to p_a is *not* a canonical transformation, and this accounts for the nonstandard form of the Poisson bracket given by Eq. (3.9), and in the modification of the usual canonical equations of motion given by Eqs. (3.10).

The geometrical significance of m_a is easily described. Let us define quantities φ_α^a by $(\partial\Pi^a/\partial z^\alpha)\varphi_\beta^\alpha = \delta_\beta^\alpha$, $\varphi_{,\alpha}\varphi_a^\alpha = 0$, and note that $(\partial\Pi^a/\partial z^\alpha)\varphi_a^\beta = \delta_\alpha^\beta - A^\beta\varphi_{,\alpha}$. Then it is not difficult to derive

$$m_a \equiv p_a - e\phi_a = \varphi_a^\alpha P_\alpha.$$

Thus at any point of Z, m_a is the (nonunique) projection of P_α onto the hypersurface through that point of a (nonunique) sectioning of Z, in contrast to p_a which is the (unique) projection of P_α orthogonal to the fiber of Z through the point in question. We see that the gauge invariant 4-momentum p_a is more natural, geometrically, than the gauge-dependent 4-momentum m_a. This provides a strong argument in favor of the nonstandard Poisson bracket [Eq. (3.9)] and the corresponding modified canonical equations of motion [Eqs. (3.10)].

C. The Hamilton-Jacobi equation

For completeness the five-dimensional Hamilton–Jacobi theory is given here, and its four-dimensional implications are derived. The procedure is to introduce Hamilton's principal function, $S(\sigma, z^\alpha)$, satisfying the partial differential equation $(\partial S/\partial\sigma) + H = 0$ and the constraints, with P_α replaced by $\partial S/\partial z^\alpha$. The constraints become

$$[(g^{\alpha\beta} - A^\alpha A^\beta)S_{,\alpha}S_{,\beta}]^{1/2} = m \tag{3.14}$$

and

$$A^\alpha S_{,\alpha} = e, \tag{3.15}$$

and modulo these equations $(\partial S/\partial\sigma) + H = 0$ becomes $\partial S/\partial\sigma = 0$. Thus the Hamilton–Jacobi theory is represented in this case by Eqs. (3.14) and (3.15), where S is independent of σ. The second constraint, Eq. (3.15), is satisfied by

$$S(z^\alpha) = S_0(z^\alpha) + e\varphi(z^\alpha), \tag{3.16}$$

if

$$A^\alpha\varphi_{,\alpha} = 1 \tag{3.17}$$

and

$$A^\alpha S_{0,\alpha} = 0. \tag{3.18}$$

The first equation shows that the function $\varphi(z^\alpha)$ represents a particular sectioning of Z, and therefore it is appropriate to denote it $\varphi(z^\alpha)$. The second equation implies that

$$S_0(z^\alpha) = S_0(\Pi^a(z^\alpha)) = S_0(x^a).$$

Substituting for S in Eq. (3.14) by means of Eq. (3.16) yields

$$[g^{ab}(S_{0,a} - e\phi_a)(S_{0,b} - e\phi_b)]^{1/2} - m = 0, \tag{3.19}$$

where $\phi_a = -A_a^\alpha\varphi_{,\alpha}$ is as usual the (nonunique) four-dimensional vector potential. Equation (3.19) is the usual relativisitc Hamilton–Jacobi equation inluding electromagnetic effects.[13]

4. QUANTUM DYNAMICS OF CHARGED PARTICLES

We wish to quantize the classical dynamical system presented in the preceding section. It is instructive to proceed in two different ways, which involve two different types of nonstandard features, and both of which yield the usual quantum mechanics of charged particles. On the one hand we can quantize the five-dimensional theory presented in Sec. III A, and then reduce this five-dimensional quantum theory to a four-dimensional one; this approach requires us to quantize a dynamics with constraints. On the other hand we can take the four-dimensional classical theory obtained in Sec. 3B by reducing the five-dimensional classical theory, and quantize it to obtain a four-dimensional quantum mechanics directly; in this case we must quantize a dynamics with a nonstandard bracket, that given by Eq. (3.9).

There is a technical question which comes up in both of the alternatives described above: what is the correct quantization procedure in a curved space? The question has not been treated in a definitive way in the literature, and a discussion is included here to justify the procedure we shall follow. In quantizing a dynamical system it is a general rule that the momentum operator, $\hat{P}_\alpha$, satisfies

$$[\hat{\mathbf{T}}, \hat{P}_\alpha] = \frac{\partial \hat{\mathbf{T}}}{\partial \hat{z}^\alpha}, \tag{4.1}$$

for $\hat{\mathbf{T}}$ a tensor function of the position operators (with indices suppressed).

We wish to justify the replacement of Eq. (4.1) by

$$[\hat{\mathbf{T}}, \hat{P}_\alpha] = \nabla_\alpha \hat{\mathbf{T}}. \tag{4.2}$$

The argument runs as follows: The mathematical analog of the result of a measurement of a physical quantity by a particular observer is always a scalar. The scalar may arise by contractions of the tensors in terms of which the theory is formulated, or by their projection onto vectors characteristic of the observer's frame of reference. Thus in principal one need deal only with scalar operators, and, in particular, one need calculate only commutators of scalars. Let us evaluate such a scalar using Eq. (4.1), and again using Eq. (4.2). If we take as our example $[\hat{A}^\alpha \hat{P}_\alpha, \hat{B}_\beta \hat{C}^\beta]$ we get, using Eq. (4.1) $\hat{A}^\alpha [\partial (B_\beta C^\beta)/\partial z^\alpha]$. If we use instead Eq. (4.2) we obtain $\hat{A}^\alpha \nabla_\alpha (\hat{B}_\beta \hat{C}^\beta)$ which is of course the same as $\hat{A}^\alpha [\partial (\hat{B}_\beta \hat{C}^\beta)/\partial z^\alpha]$. Thus even in a curved space the result is the same whether we use Eq. (4.1) or Eq. (4.2). It is easy to see that the above argument is independent of the particular example chosen. Thus we are free to use Eq. (4.2) in the evaluation of quantum brackets, and, since it results in manifest covariance at every stage of a calculation, we shall do so. Similar remarks apply to the representation of $\hat{P}_\alpha$ in the Schrödinger representation, and we shall use $-i\hbar\nabla_\alpha$ there as well.

A. Quantizing the five-dimensional theory

To quantize a Hamiltonian dynamical system with constraints we follow Dirac.[6] The first step is to replace the classical position and momentum, z^α and P_α, by operators $\hat{z}^\alpha$ and $\hat{P}_\alpha$, and postulate the usual fundamental quantum conditions

$$[\hat{z}^\alpha, \hat{z}^\beta] = 0, \quad [\hat{P}_\alpha, \hat{P}_\beta] = 0, \quad [\hat{z}^\alpha, \hat{P}_\beta] = \delta^\alpha_\beta.$$

The classical functions $g_{\alpha\beta}$ and A_α are replaced by the operators $\hat{g}_{\alpha\beta}$ and $\hat{A}_\alpha$ satisfying

$$[\hat{A}_\alpha, \hat{P}_\beta] + [\hat{A}_\beta, \hat{P}_\alpha] = 0, \tag{4.3}$$

$$\hat{A}^\alpha [\hat{A}_\alpha, \hat{P}_\beta] = 0, \tag{4.4}$$

$$\hat{A}^\alpha [\hat{A}_\beta, \hat{P}_\alpha] = 0, \tag{4.5}$$

and

$$[\hat{g}_{\alpha\beta}, \hat{P}_\gamma] = 0. \tag{4.6}$$

Due to Eqs. (2.7), (2.13), and (2.14), and Eq. (4.2), and using these equations it is easy to show that

$$[\hat{A}^\alpha, \hat{P}_\alpha] = 0. \tag{4.7}$$

If the classical constraints are replaced by

$$\hat{\Phi}_1 = \hat{A}^\alpha \hat{P}_\alpha - e \tag{4.8}$$

and[14]

$$\hat{\Phi}_2 = (\hat{g}^{\alpha\beta} - \hat{A}^\alpha \hat{A}^\beta) \hat{P}_\alpha \hat{P}_\beta - m^2, \tag{4.9}$$

one can easily show that these operators are hermitian, using Eqs. (4.3)–(4.7), and so we take as our quantum-mechanical Hamiltonian

$$\hat{H} = c_1 \hat{\Phi}_1 + c_2 \hat{\Phi}_2. \tag{4.10}$$

The constraints are certainly consistent with one another if they commute, and we have $[\hat{\Phi}_1, \hat{\Phi}_2] = [\hat{A}^\alpha \hat{P}_\alpha, (\hat{g}^{\beta\gamma} - \hat{A}^\beta \hat{A}^\gamma) \hat{P}_\beta \hat{P}_\gamma] = (\hat{g}^{\beta\gamma} - \hat{A}^\beta \hat{A}^\gamma)[\hat{A}^\alpha, \hat{P}_\beta \hat{P}_\gamma] \hat{P}_\alpha = \hat{g}^{\beta\gamma}[\hat{A}^\alpha, \hat{P}_\beta \hat{P}_\gamma] \hat{P}_\alpha = \hat{g}^{\beta\gamma} \hat{A}^\alpha_{;\gamma} \hat{P}_\beta \hat{P}_\alpha + \hat{P}_\beta \hat{A}^\alpha_{;\gamma} \hat{P}_\alpha)$, where Eqs. (4.3)–(4.7) have been used. Continuing with the aid of these equations we have $[\hat{\Phi}_1, \hat{\Phi}_2] = \hat{g}^{\beta\gamma}(\hat{A}^\alpha{}_{;\beta} \hat{P}_\gamma \hat{P}_\alpha + \hat{A}^\alpha{}_{;\gamma} \hat{P}_\beta \hat{P}_\alpha - \hat{A}^\alpha{}_{;\alpha\beta} \hat{P}_\alpha) = -\hat{g}^{\beta\gamma} \hat{A}^\alpha{}_{;\gamma\beta} \hat{P}_\alpha = -\frac{1}{2} \hat{F}^{\alpha\beta}{}_{;\beta} \hat{P}_\alpha$. It can be shown that the five-dimensional field equations, which are equivalent to the usual Einstein and Maxwell field equations, imply that $\hat{F}^{\alpha\beta}{}_{;\beta} = 0$ in vacuum.[15] Thus by assuming a source-free region of space, we avoid the complications of secondary constraints. In the Schrödinger representation the equations of motion for a system with the Hamiltonian given by Eq. (4.10) and the constraints given by Eqs. (4.8) and (4.9) are[16]

$$(-i\hbar A^\alpha \nabla_\alpha - e) = 0 \tag{4.11}$$

and

$$[(g^{\alpha\beta} - A^\alpha A^\beta)(-i\hbar \nabla_\alpha)(-i\hbar \nabla_\beta) - m^2]\Psi = 0. \tag{4.12}$$

The first two equations are solved by $\Psi(z^\alpha) = \Psi(z^\alpha) e^{ie/\hbar} \varphi(z^\alpha)$, where $A^\alpha \nabla_\alpha \psi = 0$ and $A^\alpha \nabla_\alpha \varphi = 1$. The first condition implies that $\psi(z^\alpha) = \psi(\Pi^a(z^\alpha)) = \psi(x^a)$, while the second is the only restriction on $\varphi(z^\alpha)$, so this function represents an arbitrary sectioning of Z as indicated by the notation. Substituting $\psi(z^\alpha)$ into Eq. (4.12) and writing $A^\alpha_a A_b g^{ab}$ for $g^{\alpha\beta} - A^\alpha A^\beta$, we obtain

$$-\hbar^2 g^{ab}[\nabla_a - (ie/\hbar)\phi_a][\nabla_b - (ie/\hbar)\phi_b]\psi = m^2\psi, \tag{4.13}$$

where $\phi_a \equiv -A^\alpha_a \varphi_{,\alpha}$ has been inserted. Thus we have obtained the Klein–Gordon equation modified for a charged particle.

B. Quantizing the four-dimensional theory

In Sec. 3B the four-dimensional reduction of the five-dimensional classical theory was found to be given by the equations

$$\frac{\partial x^a}{\partial \tau} = [x^a, \Phi_2], \quad \frac{\partial p_a}{\partial \tau} = [p_a, \Phi_2],$$

where the bracket is the nonstandard bracket

$$[A,B]=\frac{\partial A}{\partial x^a}\frac{\partial B}{\partial p_a}-\frac{\partial A}{\partial p_a}\frac{\partial B}{\partial x^a}+e\frac{\partial B}{\partial p_a}f_{ab}\frac{\partial A}{\partial p_b}.$$

It is obvious that this bracket is antisymmetric and linear in each argument; one can show in addition that it satisfies $[AB,C]=A[B,C]+[A,C]B$, $[A,BC]=[A,B]C+B[A,C]$, and $[A,[B,C]]+[B,[C,A]]+[C,[A,B]]=0$. Since these are all the formal properties which the quantum-mechanical bracket should have,[17] we expect that a successful quantization procedure can be based upon this bracket. Thus we replace the classical variables x^a, p_a, by the quantum mechanical operators $\hat{x}^a, \hat{p}_a$, satisfying

$$[\hat{x}^a,\hat{x}^b]=0, \quad [\hat{x}^a,\hat{p}_b]=\delta^a_b \tag{4.14}$$

$$[\hat{p}_a,\hat{p}_b]=-e\hat{f}_{ab}. \tag{4.15}$$

The existence of a Hilbert space representation of a canonical quantization scheme is usually demonstrated by the construction of the Schrödinger representation. Due to the nonstandard form of the fundamental quantum conditions it is not obvious that such a representation exists; however, a modification of the usual construction does in fact yield one and the modified procedure is considered in detail here.

We assume that the $\hat{x}^a$ for a complete set of observables for the system, and represent a state by the wave function $\psi(x^a)$. The operators x^a and $\partial/\partial x^a$ obviously satisfy $(\partial/\partial x^a)x^b\psi - x^b(\partial/\partial x^a)\psi = \delta^b_a\psi$ if ψ satisfies the usual boundary conditions; therefore, the operator $p_a + i\hbar(\partial/\partial z^a)$ satisfies the second of Eqs. (4.14), and, since the x^a form a complete set, we can write $p_a + i\hbar(\partial/\partial x^a) = f_a(x^b)$, where the f_a are as yet unknown functions of x^b. Substituting $p_a = -i\hbar(\partial/\partial x^a) + f_a$ into Eq. (4.15) yields $(\partial/\partial z^a)f_b - (\partial/\partial z^b)f_a = -ef_{ab}$, so we see that $f_a = -e\phi_a + f_{,a}$, where f is an arbitrary function. It can be shown[17] that by proper choice of phase for $\psi(x^a)$ one can set $f=0$, so we end with $p_a = -i\hbar(\partial/\partial x^a) - e\phi_a$. As before in a curved space we prefer to write

$$p_a=-i\hbar\nabla_a-e\phi_a \tag{4.16}$$

as the Schrödinger representation of p_a. With Eq. (4.16) for p_a, and $H=\Phi_2=g^{ab}p_ap_b-m^2=0$, Schrödinger's equation is

$$-\hbar^2 g^{ab}[\nabla_a-(ie/\hbar)\phi_a][\nabla_b-(ie/\hbar)\phi_b]\psi=m^2\psi, \tag{4.17}$$

which is identical to Eq. (4.13).

APPENDIX

This paper is intended to be accessible to a theoretical physicist who knows general relativity and the requisite tensor analysis, but is not a specialist in these subjects. There are two concepts from tensor analysis, that of anholonomic reference frames and that of the Lie derivative, which appear in the paper, and with which such a reader may be unfamiliar. A brief introduction to these concepts is given below (Secs. A and B). In addition, a certain result involving both of the above concepts is fundamental to the paper, but is too specialized to appear in standard references. This result is stated and proved as a theorem (Sec. C). Finally, a remark is included (Sec. D) with regard to various differential operators acting on tensors expressed in one reference frame with respect to some indices, and in another with respect to others. The results needed are elementary, but it is difficult to find an explicit discussion of them in standard sources.

A. Anholonomic Reference Frames [18]

If p is a variable ranging over the points of an m dimensional manifold, P, the functions $x^\alpha(p)$, $\alpha = 1,\ldots,m$ from P to the real numbers may serve as coordinates for the manifold. If they do, then we can define m coordinate curves intersecting at a point, p_0, of the manifold by varying each of the m coordinate values while holding the others fixed, and the unit tangent vectors of these curves will form a basis (m-independent m-vectors) at the point p_0. The "components" of a given tensor with respect to the coordinate system are its projections onto this basis, and we say the tensor is given with respect to a *holonomic* reference frame.

One can, however, use *any* basis at a point to express the tensors there, and if $b^\alpha(z^\beta)$ is a choice of basis at each point, and $b_{\alpha,\beta}-b_{\beta,\alpha}\neq 0$, then there does not exist a coordinate system whose coordinate curves lead to the basis $b^\alpha(z^\beta)$ throughout the manifold. In this case we say the tensor components are expressed in a *nonholonomic* reference frame, while if we are unsure we refer to an anholonomic reference frame. More specifically, if the transformation coefficients from some holonomic reference frame to some anholonomic reference frame are A^α_h, A^h_α, where α and h refer to the holonomic and anholonomic reference frames, respectively, then the latter is holonomic if and only if

$$\Omega^h_{ji}=A^\mu_jA^\lambda_i(A^h_{\lambda,\mu}-A^h_{\mu,\lambda})=0,$$

where Ω^h_{ji} is called the *object of anholonomity*. In this paper the anholonomic reference frame introduced was related to the z^α reference frame by the transformation coefficients $A^\alpha_h=(A^\alpha_a,A^\alpha)$ and $A^h_\alpha=((\partial\Pi/\partial z^\alpha),A_\alpha)$, where $A^\alpha A_\alpha=0, A^\alpha_a(\partial\Pi^a/\partial z^\alpha)=\delta^b_a$, and it is easy to calculate

$$\Omega^a_{hi}=0, \quad \Omega^*_{hi}=A^\alpha_hA^\beta_if_{\alpha\beta}=f_{hi}. \tag{A1}$$

Since f_{hi} is, in the physical interpretation, the electromagnetic field, the anholonomic reference frame used in this paper is in fact nonholonomic.

B. The Lie derivative

Given a differentiable manifold and a vector field, V^α, the Lie derivative of a tensor field $T^\alpha{}_{\beta\ldots}$ with respect to that vector field is defined to be[18]

$$\underset{V}{\pounds}T^\alpha{}_{\beta\ldots}=T^\alpha{}_{\beta\ldots,\gamma}V^\gamma-T^\gamma{}_{\beta\ldots}V^\alpha{}_{,\gamma}\ldots+T^\alpha{}_{\gamma\ldots}V^\gamma{}_{,\beta}+\ldots,$$

in a holonomic reference frame.

If the manifold possesses Riemannian structure this can be written

$$\underset{V}{\pounds}T^\alpha{}_{\beta\ldots}=T^\alpha{}_{\beta\ldots;\gamma}T^\gamma-T^\alpha{}_{\beta\ldots}V^\alpha{}_{;\gamma}-\ldots+T^\alpha{}_{\gamma\ldots}V^\gamma{}_{;\beta}+\ldots.$$

The Lie derivative is defined in such a way as to measure the rate of change of the tensor, $T^\alpha{}_{\beta\ldots}$, in the direction of the vector field, V^α, with the effects of the variations in the coordinatization of the manifold eliminated; thus, for example, the equation $\underset{V}{\pounds}g_{\alpha\beta}$ expresses the "constancy" of the Riemannian structure of a manifold in the direction of V^α. In the structure introduced

in Sec. 1 of this paper we have symmetry of the Riemannian structure along the fibers, and this is expressed by $\underset{A}{\pounds} g_{\alpha\beta} = 0$ [Eq. (2.7)]. In an anholonomic reference frame the definition of the Lie derivative generalizes to

$$\underset{V}{\pounds} T^h{}_{i\cdots} = T^h{}_{i\cdots,j} V^j - (V^h{}_{,j} - 2V^k \Omega^h_{jk}) T^j{}_{i\cdots} - \cdots + (V^j{}_{,i} - 2V^k \Omega^j_{ik}) T^h{}_{j\cdots} + \cdots, \quad \text{(A2)}$$

where by definition the anholonomic differentiation index, i is $A_i^\alpha(\partial/\partial z^\alpha)$.

C. A theorem

The following theorem is used in constructing tensor fields defined on X from tensor fields defined on Z: If we have

$$T^{a_1\cdots a_r}_{b_1\cdots b_s} = \frac{\partial \Pi^{a_1}}{\partial z^{\alpha_1}} \cdots \frac{\partial \Pi^{a_r}}{\partial z^{\alpha_r}} A^{\beta_1}_{b_1} \cdots A^{\beta_r}_{b_r} T^{\alpha_1\cdots\alpha_r}_{\beta_1\cdots\beta_r}$$

defined on Z, and A^α is the unit vector field tangent to the fibers of Z, then

$$\underset{A}{\pounds} T^{a_1\cdots a_r}_{b_1\cdots b_s} = T^{a_1\cdots a_r}_{b_1\cdots b_s,\gamma} A^\gamma. \quad \text{(A3)}$$

Restricting the free indices in Eq. (A2) to the range 0, 1, 2, 3 we have

$$\underset{A}{\pounds} T^{b_1\cdots b_r}_{a_1\cdots a_s} = T^{b_1\cdots b_r}_{a_1\cdots a_s,k} A^k + T^{b_1\cdots b_r}_{k a_2\cdots a_s} (A^k{}_{,a_1} - 2A^i \Omega^k_{a_1 i}) + \cdots - T^{k b_2\cdots b_r}_{a_1\cdots a_s} A^{b_1}{}_{,k} - 2A^i \Omega^{b_1}_{k i}) - \cdots$$

where the dummy indices i and k necessarily retain the range $-1, 0, 1, 2, 3$. By definition we have $A^k = A^k_\alpha A^\alpha = ((\partial \Pi^a/\partial z^\alpha) A^\alpha, A_\alpha A^\alpha) = (1, 0, 0, 0, 0)$, so $A^k{}_{,a_1} = 0$ and $A^{b_1}{}_{,k} = 0$, while $A^i \Omega^{b_1}_{k i} = 0$ since $\Omega^b_{ki} = 0$ [Eq. (A1)]. Thus we are left with

$$A^{b_1\cdots b_r}_{a_1\cdots a_s} = T^{b_1\cdots b_r}_{a_1\cdots a_s,k} A^k - 2T^{b_1\cdots b_r}_{k a_2\cdots a_s} A^i \Omega^k_{a_1 i} - \cdots.$$

Since $A^k = (1, 0, 0, 0, 0)$ we see that $A^i \Omega^k_{a_1 i} = (\Omega^\bullet_{a_1 \cdot}, 0)$ [Eq. (A1) again], and $\Omega^\bullet_{a_1 \gamma} = A^\alpha_{a_1} A^\beta (\nabla_\alpha A_\beta - \nabla_\beta A_\alpha) = 0$ in view of Eqs. (2.7) and (2.13), giving the desired result.

D. "Mixed" tensors

If one wishes to calculate the covariant derivative of a tensor expressed simultaneously in two different reference frames, for example $T^{(\alpha')}_\beta = (\partial x^{\alpha'}/\partial x^\gamma) T^\alpha_\beta$, this reduces to the same operator on the "mixed" identity, $\partial x^{\alpha'}/\partial x^\gamma$. Thus we have

$$\nabla_\gamma T^{(\alpha')}_\beta = \left(\nabla_\gamma \frac{\partial x^{\alpha'}}{\partial x^\gamma}\right) T^\gamma_\beta + \frac{\partial x^{\alpha'}}{\partial x^\gamma} \nabla_\gamma T^\gamma_\beta,$$

and $\dfrac{\partial x^{\alpha'}}{\partial x^\gamma}$ is the "mixed" identity. However,

$$\nabla_\gamma \frac{\partial x^{\alpha'}}{\partial x^\beta} = \frac{\partial^2 x^{\alpha'}}{\partial x^\gamma \partial x^\beta} + \left\{{}^{\delta}_{\gamma\beta}\right\} \frac{\partial x^{\delta}}{\partial x^\gamma} - \left\{{}^{\delta}_{\gamma\beta}\right\} \frac{\partial x^{\alpha'}}{\partial x^\delta}$$

which vanishes by virtue of the transformation law of Christoffel symbols. Similarly it can be shown that

$$\underset{V}{\pounds} \left(\frac{\partial x^{\alpha'}}{\partial x^\beta}\right) = 0.$$

If follows directly that the Lie and covariant derivatives of the "mixed" metric tensor vanish.

[1] T. Kaluza, Sitzungsber. Preuss. Akad. Wiss. Phys.-Math. Kl. (1921). (1921).

[2] P. G. Bergmann, *Introduction to the Theory of Relativity* (Prentice-Hall, Englewood Cliffs, N. J., 1942).

[3] W. Tulczyjew, Zeszyty Nauk. Uniw. Jagiel. Mat. 3, 143 (1965).

[4] P. G. Bergmann and J. L. Anderson, Phys. Rev. **83**, 1018 (1951).

[5] P. Dirac, Can. J. Math. **2**, 129 (1950).

[6] P. Dirac, *Lectures on Quantum Mechanics* (Belfer Graduate School of Science, Yeshiva University, New York, 1964).

[7] Lower case Greek and Latin indices have the ranges 0, 1, 2, 3, 4 and 1, 2, 3, 4, respectively, and indicate holonomic reference frames, except that h, i, $\cdots$ will have the range 0, 1, 2, 3, 4 and indicate an anholonomic reference frame. (The 0 component in an anholonomic reference frame may also be denoted by $\cdot$, that is, $\xi^0 \equiv \xi\bullet$.) Partial differentiation will be written out or indicated by a comma, while covariant differentiation will be denoted by the symbol ∇_a or indicated by a semicolon. Quantum mechanical operators will be distinguished by $\hat{}$, that is, $\hat{P}_a$ is the operator corresponding to the classical variable P_a.

[8] The functions $\Pi(z^a)$ and $A(z^a)$ must be assumed differentiable (see Eqs. 2.5 and 2.6).

[9] In the interest of logical economy one can identify the one-dimensional manifold and the additive group of the real numbers γ which parametrize the group of motions.

[10] The Lie derivative is defined in the Appendix for those unfamiliar with it.

[11] Anholonimic reference frames are discussed in the Appendix.

[12] A note on various differential operations applied to the "mixed identity" is included in the Appendix.

[13] L. Landan and E. Lifshitz, *The Classical Theory of Fields* (Pergamon, New York, 1962).

[14] The constraint $(g^{ab} p_a p_b)^{(1/2)} = m$ has been replaced by its square $g^{ab} p_a p_b = m^2$. It is still designated by $\Phi_2 = 0$. As long as m is assumed to be positive the two forms are equivalent.

[15] This result is not difficult to obtain using equations derived in Ref. 3 and an adapted coordinate system.[2]

[16] It is known[6] that the absence of an absolute time variable corresponds to a Lagrangian homogenous in velocities and a Hamiltonian constrained to vanish. It is obvious that under these circumstances Hamilton's principal function and the wavefunction of the Schrödinger representation will be independent of time.

[17] P. Dirac, *The Principles of Quantum Mechanics* (Oxford U. P., Toronto, 1958).

[18] Schouten, *Ricci Calculus* (Springer, Berlin, 1954).

Power statistics for wave propagation in one dimension and comparison with radiative transport theory*

W. Kohler† and G. C. Papanicolaou

Courant Institute of Mathematical Sciences, New York University, New York, New York
(Received 11 May 1973)

We consider a one-dimensional medium with random index of refraction or a transmission line with random capacitance per unit length, allowing for impedance mismatch at the load and generator. We compute the expected value of the incident and reflected powers at any point between the generator and load in the limit of weak fluctuations and a long line. The results are compared with those of radiative transport theory and discrepancies show the limitations of that theory. Finally, we consider the spreading of pulses due to random fluctuations.

1. INTRODUCTION

Consider a transmission line with constant inductance per unit length and capacitance per unit length which is a random function of position, fluctuating slightly from a constant expected value. The current and voltage, or the forward and backward travelling waves, are random functions of position. We wish to determine their statistical characteristics at any point between the load and generator when the line is long and the fluctuations are weak. This transmission line formulation is applicable to one-dimensional wave propagation in a random medium and the propagation of the fundamental mode in a waveguide with random inhomogeneities. These problems all lead to the same mathematical questions.

Using a general limit theorem[1,2] we compute the mean of the forward and backward propagating powers in the interior of the line for a very broad class of fluctuation processes. We also allow for passive impedance mismatch at the generator and the load. For the one-dimensional random medium the index of refraction in the absence of fluctuations is allowed to take different values in the interior and two exterior regions. When the uniform (no fluctuations) line is matched to the generator and load the above results have been obtained previously for two particular choices of fluctuation processes, one by Gazaryan[3] and one by Lang.[4]

We proceed next to compare these results to those obtained by using radiative transport theory.[5,6] Transport theory has been used to study wave propagation in waveguides and random media by Marcuse[7] and others.[8-14] The comparison shows that transport theory does not account properly for the behavior of the fields in the interior. The absolute error is not large but the relative error is often quite large (cf. Figs. 2-4, 6). Transport theory, as well as the analysis based on the limit theorem, is supposed to hold under the following general conditions: weak fluctuations, long line, and short correlation length for the fluctuations. We are led therefore to the conclusion that transport theory is not valid without additional restrictions. When the line is matched to the generator and load the two theories are in good agreement if we restrict a dimensionless line length parameter to be small. In the mismatched case the discrepancies are more pronounced as one would expect from physical considerations.

In Sec. 2 we formulate the problem under consideration both for transmission lines and one-dimensional random media and reduce it to the analysis of a stochastic two point boundary value problem. In Sec. 3 we express the solution of the boundary value problem in terms of fundamental matrices or propagators. In Sec. 4 we state the limit theorem mentioned above in the form of Ref. 2 and use it for the propagator matrices. In Sec. 5 we introduce a convenient parametrization for the matrices. Section 6 contains our main result stated as a theorem, which generalizes an observation of Gazaryan.[3] We obtain a simple heat equation for the expectation of the total power in the interior where the length of the line plays the role of time and distance from the middle of the line the space variable. In Sec. 7 we recover results of J. A. Morrison[15] concerning the mean of the modulus square of the transmission coefficient. Since the latter depends only on the length of the line, it may be called an exterior field quantity.

Section 8 contains a summary of transport theory for the present problem. Here we display graphs that show the discrepancies mentioned above.

Finally, in Sec. 9 we treat the problem of pulse spreading as a result of the fluctuations. This is conveniently done here since we can use the machinery that has been set up in previous sections.

A general survey of work on wave propagation in random media is given by Frisch.[16] More recent work is presented by Morrison and McKenna.[17]

2. FORMULATION OF THE PROBLEM

We begin with a transmission line formulation and then consider wave propagation in a one-dimensional random medium. Both problems lead to the stochastic two point boundary value problem (2.6), (2.7).

Let $V(x)$ and $I(x)$ denote the complex-valued voltage and current at x on a transmission line which occupies the interval $0 \le x \le l$. The time factor $e^{-i\omega t}$ will be omitted throughout. V and I satisfy the boundary value problem

$$\frac{dV(x)}{dx} = i\omega L(x)I(x),$$

$$\frac{dI(x)}{dx} = i\omega C(x)V(x), \qquad 0 \le x \le l, \tag{2.1}$$

$$e_g - V(0) = I(0)Z_g, \qquad V(l) = I(l)Z_l. \tag{2.2}$$

Here $L(x)$ and $C(x)$ are the inductance and capacitance per unit length, $Z_g = Z_g(\omega)$ and $Z_l(\omega)$ are the generator and load impedances and $e_g = e_g(\omega)$ is the generator voltage (cf. Fig. 1).

We wish to study (2.1) and (2.2) when $L(x) = L_0$ is constant and $C(x)$ is a random function of x which fluctuates slightly from its constant expected value C_0. Thus we let

$$C(x) = C_0 + \epsilon C_1(x), \qquad E\{C_1(x)\} = 0. \tag{2.3}$$

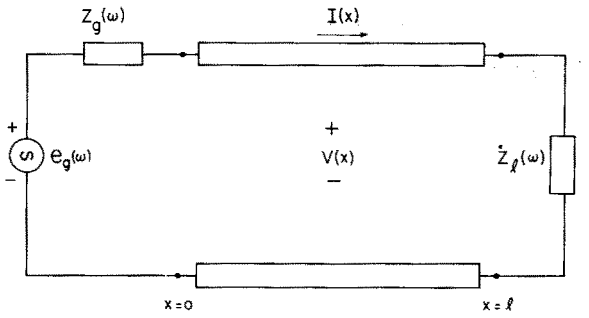


FIG. 1.

We use the notation $E\{\cdot\}$ for the operation of taking expected values and denote by ϵ a small real parameter characterising the size of the fluctuations. Other properties of $C_1(x)$ will be specified later.

Let us introduce the characteristic impedance Z_0, admittance Y_0, the speed of propagation c, and the wave number k of the uniform ($\epsilon = 0$) line.

$$Z_0 = Y_0^{-1} = (L_0/C_0)^{1/2}, \quad c = (L_0C_0)^{-1/2}, \quad k = \omega/c. \tag{2.4}$$

Let us also define forward and backward traveling wave amplitudes $A(x)$ and $B(x)$, which we expect to be "slowly varying," by

$$\begin{aligned} A(x) &= \tfrac{1}{2}e^{-ikx}[Y_0^{1/2}V(x) + Z_0^{1/2}I(x)], \\ B(x) &= \tfrac{1}{2}e^{ikx}[Y_0^{1/2}V(x) - Z_0^{1/2}I(x)]. \end{aligned} \tag{2.5}$$

On using (2.1), (2.2), and (2.3) it follows that $A(x)$ and $B(x)$ satisfy the stochastic two point boundary value problem

$$\begin{aligned} \frac{dA(x)}{dx} &= \epsilon\frac{ik\mu(x)}{2}[A(x) + B(x)e^{-2ikx}], \\ \frac{dB(x)}{dx} &= -\epsilon\frac{ik\mu(x)}{2}[A(x)e^{2ikx} + B(x)], \quad 0 \le x \le l, \end{aligned} \tag{2.6}$$

$$A(0) = E_g + \Gamma_g B(0), \quad B(l) = \Gamma_l A(l). \tag{2.7}$$

Here we have introduced the notation

$$\mu(x) = C_1(x)/C_0, \quad E_g = e_g Z_0^{1/2}/(Z_0 + Z_g) \tag{2.8}$$

$$\begin{aligned} \Gamma_g &= (Z_g - Z_0)/(Z_g + Z_0), \\ \Gamma_l &= e^{2ikl}[(Z_l - Z_0)/(Z_l + Z_0)]. \end{aligned} \tag{2.9}$$

The quantities Γ_g and Γ_l are generator and load reflection coefficients for the uniform ($\epsilon = 0$) line and E_g is a normalized generator output. When $Z_g = Z_0$ then $\Gamma_g = 0$ and the uniform line is matched to the generator while when $Z_l = Z_0$ then $\Gamma_l = 0$ and it is matched to the load. Note that A and B are complex functions of x, $0 \le x \le l$, the length of the line l and the wave number k. The dependence on k will not be displayed until Sec. 9.

Consider next a one-dimensional random medium occupying the interval $0 \le x \le l$. Let $u(x)$ and $n(x)$ be the wave field ($e^{-i\omega t}$ omitted) and the index of refraction at location x, respectively. We assume that $u(x)$, $-\infty < x < \infty$, satisfies the reduced wave equation

$$\frac{d^2u(x)}{dx^2} + k^2n^2(x)u(x) = 0, \quad -\infty < x < \infty, \tag{2.10}$$

$$n^2(x) = \begin{cases} n_1^2, & x < 0, \\ 1 + \epsilon\mu(x), & 0 \le x \le l, \\ n_2^2, & x > l, \end{cases} \tag{2.11}$$

$$u(x) \text{ and } \frac{du(x)}{dx} \text{ continuous.} \tag{2.12}$$

As before, $\mu(x)$ denotes a zero mean random process. The problem (2.10)-(2.12) is completed by specifying that a plane wave of unit amplitude is incident from the left. If we denote by R and T the complex-valued reflection and transmission coefficients, then we have

$$\begin{aligned} u(x) &= e^{ikn_1x} + Re^{-ikn_1x}, \quad x < 0, \\ u(x) &= Te^{ikn_2x}, \quad x > l. \end{aligned} \tag{2.13}$$

From (2.13) and (2.10)-(2.12) we obtain the following stochastic two point boundary value problem for $u(x)$:

$$\begin{aligned} &\frac{d^2u(x)}{dx^2} + k^2[1 + \epsilon\mu(x)]u(x) = 0, \quad 0 \le x \le l, \\ &\tfrac{1}{2}\left[u(0) + \frac{1}{ikn_1}\frac{du(0)}{dx}\right] = 1, \quad \frac{du(l)}{dx} = ikn_2u(l). \end{aligned} \tag{2.14}$$

Finally, we let

$$\begin{aligned} u(x) &= e^{ikx}A(x) + e^{-ikx}B(x) \\ \frac{du(x)}{dx} &= ik[e^{ikx}A(x) - e^{-ikx}B(x)], \quad 0 \le x \le l, \end{aligned} \tag{2.15}$$

and deduce from (2.14) that $A(x)$ and $B(x)$ satisfy the two point boundary value problem (2.6), (2.7). Instead of (2.8), (2.9) we have now

$$E_g = 2n_1/(1 + n_1), \tag{2.16}$$

$$\Gamma_g = (1 - n_1)/(1 + n_1), \quad \Gamma_l = e^{2ikl}[(1 - n_2)/(1 + n_2)]. \tag{2.17}$$

The goal in investigating (2.6), (2.7) is to compute the expected value of $|A|^2$ and $|B|^2$, the incident and reflected power, respectively, as functions of x, $0 \le x \le l$, with l variable, $l \ge 0$. We shall do this asymptotically when ϵ is small and l is large in a manner which we specify in Sec. 4.

Since the line is lossless, the power flux

$$\mathrm{Re}\{V\bar{I}\} = |A|^2 - |B|^2, \tag{2.18}$$

is independent of x, $0 \le x \le l$. In (2.18) the bar denotes complex conjugate. In view of the conservation law (2.18), instead of computing the expectations of $|A|^2$ and $|B|^2$ directly, we will compute the expectations of $|A|^2 + |B|^2$ and $|A|^2 - |B|^2$. This is done in Secs. 6 and 7, respectively.

3. PROPAGATOR MATRICES

In this section we express the solution of the stochastic boundary value problem (2.6), (2.7) in terms of propagator or fundamental solution matrices. This is convenient because the statistical properties of the latter can be obtained using limit theorems as will be shown in the next section.

Let $m(x)$ be the 2×2 matrix valued stochastic process defined by

$$m(x) = k\mu(x)\begin{pmatrix} i/2 & (i/2)e^{-2ikx} \\ -(i/2)e^{2ikx} & -i/2 \end{pmatrix}. \tag{3.1}$$

Note that $m(x) = m(x;k)$ but, until Sec. 9, we shall not

show the dependence on k explicitly. Let $Y(x)$ be any matrix valued solution of

$$\frac{dY(x)}{dx} = \epsilon m(x)Y(x).$$

Since $\mathrm{Tr}m(x) = 0$, it follows that $\det Y(x) = \text{const}$ and we take this constant to equal 1. Furthermore, because of the form of $m(x)$ in (3.1) $Y = Y(x)$ has the form

$$Y = \begin{pmatrix} a & b \\ \bar{b} & \bar{a} \end{pmatrix}, \quad |a|^2 - |b|^2 = 1. \tag{3.2}$$

The collection of all such matrices forms a group which is denoted by $SU(1,1)$.[18]

Because the generator and load impedances are passive the reflection coefficients Γ_g and Γ_l are complex numbers of modulus less than or equal to one. Hence we may write

$$\Gamma_g = b_g/\bar{a}_g, \quad |a_g|^2 - |b_g|^2 = 1, \tag{3.3}$$

$$\Gamma_l = -\bar{b}_l/\bar{a}_l, \quad |a_l|^2 - |b_l|^2 = 1. \tag{3.4}$$

It is clear that a_g, b_g and a_l, b_l are not determined uniquely by (3.3) and (3.4). This, however, does not affect the results below. In view of (3.3) and (3.4) it is convenient to introduce generator and load matrices Y_g and Y_l

$$Y_g = \begin{pmatrix} a_g & b_g \\ \bar{b}_g & \bar{a}_g \end{pmatrix}, \quad Y_l = \begin{pmatrix} a_l & b_l \\ \bar{b}_l & \bar{a}_l \end{pmatrix}. \tag{3.5}$$

These are constant matrices belonging to $SU(1,1)$.

Consider now the solution matrices $Y_1(x,0)$ and $Y_2(l,x)$ of the following initial and final value problems:

$$\frac{dY_1(x,0)}{dx} = \epsilon m(x)Y_1(x,0), \quad x \geq 0, \quad Y_1(0,0) = Y_g, \tag{3.6}$$

$$\frac{dY_2(l,x)}{dx} = -\epsilon Y_2(l,x)m(x), \quad x \leq l, \quad Y_2(l,l) = Y_l. \tag{3.7}$$

From the remarks above it follows that Y_1 and Y_2 are stochastic processes with values in $SU(1,1)$. Moreover, it can be verified by direct computation that $Y_2(l,x)Y_1(x,0) \in SU(1,1)$ is independent of x. For Y_1 and Y_2 we use the notation

$$Y_i = \begin{pmatrix} a_i & b_i \\ \bar{b}_i & \bar{a}_i \end{pmatrix}, \quad |a_i|^2 - |b_i|^2 = 1, \quad i = 1,2, \tag{3.8}$$

where quantities with subscript 1 are functions of $x \geq 0$ and those with subscript 2 are functions of l and x, $0 \leq x \leq l$.

We now define $A(l,x)$ and $B(l,x)$ as follows:

$$A(l,x)/\bar{a}_g E_g = \bar{a}_2/(\bar{b}_2 b_1 + \bar{a}_2 \bar{a}_1), \tag{3.9}$$

$$B(l,x)/\bar{a}_g E_g = -\bar{b}_2/(\bar{b}_2 b_1 + \bar{a}_2 \bar{a}_1), \quad 0 \leq x \leq l. \tag{3.10}$$

Since the denominator on the right side of (3.9) and (3.10) is the 2,2 element of $Y_2(l,x)Y_1(x,0)$ it is independent of x as observed above. From this fact and (3.6), (3.7) it follows by direct computation that $A(l,x)$ and $B(l,x)$ of (3.9) and (3.10) satisfy the boundary value problem (2.6), (2.7). We have thus the desired expressions for $A(l,x)$ and $B(l,x)$ in terms of the propagator matrices $Y_1(x,0)$ and $Y_2(l,x)$. Note that the right sides of (3.9) and (3.10) are fixed functions of the elements of Y_1 and Y_2 and are not explicitly dependent upon the generator and load matrices Y_g and Y_l.

4. LIMIT THEOREM FOR THE PROPAGATOR MATRICES

So far we have not considered the statistical nature of the problem. In this section we study the matrix valued processes $Y_1(x,0)$, $Y_2(l,x)$ of (3.6) and (3.7) through which $A(l,x)$ and $B(l,x)$ are given by (3.9) and (3.10). Until the end of this section we consider $Y_1(x,0)$ only and so we shall drop the subscript.

We observed in Sec. 3 that $Y(x,0)$ is a process with values in $SU(1,1)$ which is a Lie group.[18] We shall denote the Lie algebra of $SU(1,1)$ by $su(1,1)$ and select the following basis in $su(1,1)$:

$$\eta_1 = \tfrac{1}{2}\begin{bmatrix} i & 0 \\ 0 & -i \end{bmatrix}, \quad \eta_2 = \tfrac{1}{2}\begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}, \quad \eta_3 = \tfrac{1}{2}\begin{bmatrix} 0 & i \\ -i & 0 \end{bmatrix}. \tag{4.1}$$

The matrices η_1, η_2, η_3 satisfy the commutation relations

$$[\eta_1, \eta_2] = \eta_3, \quad [\eta_1, \eta_3] = -\eta_2, \quad [\eta_2, \eta_3] = -\eta_1. \tag{4.2}$$

In terms of this basis the $su(1,1)$ valued stochastic process $m(x)$, defined by (3.1), has the form

$$m(x) = k\mu(x)\eta_1 + [k\mu(x)\sin 2kx]\eta_2 + [k\mu(x)\cos 2kx]\eta_3 = \sum_{j=1}^{3} m_j(x)\eta_j. \tag{4.3}$$

Let C denote the class of bounded continuous functions on $SU(1,1)$ which have a finite limit at infinity. For sufficiently smooth functions in C we define the differential operators D_{η_j}, $1 \leq j \leq 3$, by

$$(D_{\eta_j} f)(Y) = \lim_{h \downarrow 0} [f(e^{h\eta_j} Y) - f(Y)]/h. \tag{4.4}$$

Here $e^{h\eta_j}$ denotes the exponential of the matrix $h\eta_j$, $1 \leq j \leq 3$, and the limit is taken with respect to the maximum norm in C. In the next section we shall express D_{η_j} concretely in terms of coordinates parametrizing $SU(1,1)$.

We now introduce hypotheses about the random function $\mu(x)$ in (3.1), where $E\{\mu(x)\} = 0$ in view of (2.3) and (2.8). We assume that $\mu(x)$, $-\infty < x < \infty$, is a stationary random function on a probability space $(\Omega, \mathfrak{F}, \mathcal{P})$ and it is almost surely bounded, say $|\mu(x)| \leq 1$. Let $\mathfrak{F}_{x_1}^{x_2} \subset \mathfrak{F}$ denote the σ-algebra generated by $\mu(x)$, $x_1 \leq x \leq x_2$. We require that $\mu(x)$ be mixing as follows: if A is any set in $\mathfrak{F}_0^x$ and B any set in $\mathfrak{F}_{x+s}^{\infty}$, then

$$\sup_{x \geq 0} |P(B|A) - P(B)| = \rho(s) \downarrow 0, \quad s \to \infty, \tag{4.5}$$

and, in addition,

$$\int_0^\infty \rho^{1/2}(s)ds < \infty. \tag{4.6}$$

The physical meaning of (4.5) and (4.6) is that the fluctuation process $\mu(x)$ is such that $\mu(x_1)$ and $\mu(x_2)$ tend to become independent random variables sufficiently rapidly as $|x_1 - x_2| \to \infty$.

Theorem 3 of Ref. 2 can now be stated as follows. Let $\sigma = \epsilon^2 x$ and set

$$Y^{(\epsilon)}(\sigma) = Y(\sigma/\epsilon^2, 0), \tag{4.7}$$

$$u^{(\epsilon)}(\sigma, Y_g) = E\{f(Y^{(\epsilon)}(\sigma))\}, \quad f \in C. \tag{4.8}$$

Then $u^{(\epsilon)}(\sigma, Y_g)$ converges to $u^{(0)}(\sigma, Y_g)$ as $\epsilon \to 0$ and σ remains fixed, where $u^{(0)}(\sigma) = u^{(0)}(\sigma, Y_g)$ satisfies the Cauchy problem

$$\frac{\partial}{\partial \sigma} u^{(0)} = \sum_{i,j=1}^{3} a^{ij} D_{\eta_i} D_{\eta_j} u^{(0)} \equiv V u^{(0)},$$

$$\sigma > 0, \quad u^{(0)}(0, Y_g) = f(Y_g). \tag{4.9}$$

The coefficients a^{ij} are given by the formula

$$a^{ij} = \lim_{t \to \infty} \frac{1}{t} \int_{t_0}^{t} \int_{t_0}^{s} E\{m_i(\sigma) m_j(s)\} d\sigma ds, \tag{4.10}$$

which is independent of t_0 as required in Theorem 3.[2]

In view of (4.3) the a^{ij} of (4.10) can be computed explicitly. On using the fact that $D_{\eta_2} D_{\eta_3} - D_{\eta_3} D_{\eta_2} = -D_{\eta_1}$, which follows from (4.2) and (4.4), we obtain the following expression for the operator V in (4.9):

$$V = \alpha(D_{\eta_2} D_{\eta_2} + D_{\eta_3} D_{\eta_3}) + \gamma D_{\eta_1} D_{\eta_1} - \beta D_{\eta_1}, \tag{4.11}$$

$$\alpha = \frac{k^2}{2} \int_0^\infty R(s) \cos 2ks ds, \quad \beta = \frac{k^2}{2} \int_0^\infty R(s) \sin 2ks ds,$$
$$\gamma = k^2 \int_0^\infty R(s) ds, \tag{4.12}$$

$$R(s) = E\{\mu(x + s)\mu(x)\}. \tag{4.13}$$

Theorem 3 asserts,[2] in addition to the above, that if $f \in C$ is sufficiently smooth and the limit in (4.10) is approached sufficiently rapidly then the error in approximating $u^{(\epsilon)}(\sigma, Y_g)$ by $u^{(0)}(\sigma, Y_g)$ is $0(\epsilon)$.

Let $Y^{(0)}(\sigma)$, $\sigma \geq 0$, $Y^{(0)}(0) = Y_g$ be the diffusion Markov process with infinitesimal generator V given by (4.11). Since the operator V is the right invariant and independent of σ it follows that $Y^{(0)}(\sigma)$ is also a process with stationary independent increments. First, let us show that V is right invariant. Let $R_{\tilde{Y}}$, $\tilde{Y} \in SU(1,1)$, be an operator on C defined by

$$(R_{\tilde{Y}} f)(Y) = f(Y\tilde{Y}). \tag{4.14}$$

From the definitions (4.4) and (4.14), we have

$$R_{\tilde{Y}} D_{\eta_j} = D_{\eta_j} R_{\tilde{Y}}, \quad 1 \leq j \leq 3, \tag{4.15}$$

which means that the differential operators D_{η_j} commute with right translations or are right invariant. Since V in (4.11) is expressed in terms of the D_{η_j} with constant coefficients our assertion that V is right invariant follows. Now, $Y^{(0)}(\sigma)$ being a process with stationary independent increments means that if $0 \leq \sigma_1 \leq \sigma_2 \leq \cdots \leq \sigma_n < \infty$ then the random matrices

$$Y^{(0)}(\sigma_1) Y_g^{-1},$$
$$Y^{(0)}(\sigma_2)[Y^{(0)}(\sigma_1)]^{-1}, \ldots, Y^{(0)}(\sigma_n)[Y^{(0)}(\sigma_{n-1})]^{-1}, \tag{4.16}$$

are independent and their distribution depends only on the increments of the parameter: $\sigma_1, \sigma_2 - \sigma_1, \ldots, \sigma_n - \sigma_{n-1}$.

Theorem 3 of Ref. 2 shows that for a fixed σ, $Y^{(\epsilon)}(\sigma)$ converges weakly to $Y^{(0)}(\sigma)$ as $\epsilon \to 0$. The argument used there, however, can be used to show that all finite dimensional distributions of $Y^{(\epsilon)}(\sigma)$ converge to those of $Y^{(0)}(\sigma)$, $\sigma < \infty$. Weak convergence of the processes requires additional argument but we do not need it here.

Let $\tau = \epsilon^2 l$. In view of the definition (4.4) of $Y^{(\epsilon)}(\sigma)$ and (3.6), (3.7) we have, as $\epsilon \to \infty$,

$$Y_1(\sigma/\epsilon^2, 0) = Y^{(\epsilon)}(\sigma) \to Y^{(0)}(\sigma), \quad \sigma \geq 0, \tag{4.17}$$

$$Y_2(\tau/\epsilon^2, \sigma/\epsilon^2) = Y_l Y^{(\epsilon)}(\tau)[Y^{(\epsilon)}(\sigma)]^{-1}$$
$$\to Y_l Y^{(0)}(\tau)[Y^{(0)}(\sigma)]^{-1}, \quad 0 \leq \sigma \leq \tau. \tag{4.18}$$

Here the arrow denotes weak convergence and, because of the independent increments property of $Y^{(0)}$, the limit matrices corresponding to Y_1 and Y_2 are independent.

The distribution of $Y_l Y^{(0)}(\tau)[Y^{(0)}(\sigma)]^{-1}$ can be obtained in the following way, similar to the one for $Y^{(0)}(\sigma)$. Let $f \in C$ and $u^{(0)}(\tau - \sigma, Y_l) = E\{f(Y_l Y^{(0)}(\tau)[Y^{(0)}(\sigma)]^{-1})\}$. Then $u^{(0)}(\sigma, Y_l)$ satisfies the Cauchy problem

$$\frac{\partial}{\partial \sigma} u^{(0)} = \tilde{V} u^{(0)}, \quad \sigma \geq 0, \quad u^{(0)}(0, Y_l) = f(Y_l), \tag{4.19}$$

where $\tilde{V}$ is identical with V of (4.11) except the differential operators D_{η_j} are replaced by $\tilde{D}_{\eta_j}$ and

$$(\tilde{D}_{\eta_j} f)(Y) = \lim_{h \downarrow 0} \frac{f(Y e^{h\eta_j}) - f(Y)}{h}, \quad 1 \leq j \leq 3. \tag{4.20}$$

The result (4.19) is convenient because it takes care of the final value condition for Y_2 in (3.7) in the same way that the initial condition for Y_1 is taken care of in the limit by conditioning $Y^{(0)}(\sigma)$ so that $Y^{(0)}(0) = Y_g$.

5. POLAR COORDINATES

In order to implement the asymptotic results of the previous section we must introduce a convenient parametrization of $SU(1,1)$. This is done by the Euler angle or polar coordinate parametrization[19]

$$Y = \begin{pmatrix} a & b \\ \bar{b} & \bar{a} \end{pmatrix} = \begin{pmatrix} e^{i(\phi+\psi)/2} \cosh(\theta/2) & e^{i(\phi-\psi)/2} \sinh(\theta/2) \\ e^{-i(\phi-\psi)/2} \sinh(\theta/2) & e^{-i(\phi+\psi)/2} \cosh(\theta/2) \end{pmatrix}, \tag{5.1}$$

$$0 \leq \theta < \infty, \quad 0 \leq \phi < 2\pi, \quad 0 \leq \psi < 4\pi. \tag{5.2}$$

The coordinates (θ, ϕ, ψ) are global coordinates on $SU(1,1)$. Functions on $SU(1,1)$ are simply functions of (θ, ϕ, ψ). In (5.2) we could have allowed $\phi \in [0, 4\pi)$ and $\psi \in [0, 2\pi)$ and, in considerations below when one of these angles plays no role we always assume it is the one in $[0, 4\pi)$.

Let us represent the reflection coefficients (3.3), (3.4) as

$$\Gamma_g = e^{i\varphi_g} \tanh(\theta_g/2), \quad \Gamma_l = -e^{i\psi_l} \tanh(\theta_l/2). \tag{5.3}$$

From (3.5) and (5.3) the generator and load matrices become

$$Y_g = \begin{pmatrix} e^{i\varphi_g/2} \cosh(\theta_g/2) & e^{i\varphi_g/2} \sinh(\theta_g/2) \\ e^{-i\varphi_g/2} \sinh(\theta_g/2) & e^{-i\varphi_g/2} \cosh(\theta_g/2) \end{pmatrix}, \tag{5.4}$$

$$Y_l = \begin{pmatrix} e^{i\psi_l/2} \cosh(\theta_l/2) & e^{-i\psi_l/2} \sinh(\theta_l/2) \\ e^{i\psi_l/2} \sinh(\theta_l/2) & e^{-i\psi_l/2} \cosh(\theta_l/2) \end{pmatrix}. \tag{5.5}$$

Here we have resolved the multivaluedness mentioned after (3.4) by taking $\psi_g = 0$ and $\phi_l = 0$. We shall also consider ψ_l as a fixed parameter of the problem and not affected by the asymptotic limit of the previous section even though, in view of (2.9), ψ_l depends explicitly on l.

The processes $Y_1(x,0)$ and $Y_2(l,x)$ of (3.6), (3.7) are parametrized by (5.1) with subscripts 1 and 2 on (θ,ϕ,ψ). As in the previous section we introduce the notation

$$\theta_1^{(\epsilon)}(\sigma)=\theta_1(\sigma/\epsilon^2),\quad \phi_1^{(\epsilon)}(\sigma)=\phi_1(\sigma/\epsilon^2),\quad \psi_1^{(\epsilon)}(\sigma)=\psi_1(\sigma/\epsilon^2),\quad \sigma=\epsilon^2 x, \tag{5.6}$$

$$\theta_2^{(\epsilon)}(\tau,\sigma)=\theta_2(\tau/\epsilon^2,\sigma/\epsilon^2),\quad \phi_2^{(\epsilon)}(\tau,\sigma)=\phi_2(\tau/\epsilon^2,\sigma/\epsilon^2),\quad \psi_2^{(\epsilon)}(\tau,\sigma)=\psi_2(\tau/\epsilon^2,\sigma/\epsilon^2),\quad \tau=\epsilon^2 l. \tag{5.7}$$

The limit process $Y^{(0)}(\sigma)$ is parametrized by $(\theta^{(0)}(\sigma), \phi^{(0)}(\sigma), \psi^{(0)}(\sigma))$.

From their definition in (4.4) and (5.1) it follows by direct computation that the differential operators D_{η_i}, $1\le i\le 3$, are given by

$$D_{\eta_1}=\frac{\partial}{\partial\phi}, \tag{5.8}$$

$$D_{\eta_2}=-\sin\phi\coth\theta\frac{\partial}{\partial\phi}+\sin\phi\,\operatorname{csch}\theta\frac{\partial}{\partial\psi}+\cos\phi\frac{\partial}{\partial\theta}, \tag{5.9}$$

$$D_{\eta_3}=\cos\phi\coth\theta\frac{\partial}{\partial\phi}-\cos\phi\,\operatorname{csch}\theta\frac{\partial}{\partial\psi}+\sin\phi\frac{\partial}{\partial\theta}. \tag{5.10}$$

On using (5.8)-(5.10) in (4.11), V becomes

$$V=\alpha\left[\frac{\partial^2}{\partial\theta^2}+\coth\theta\frac{\partial}{\partial\theta}+\left(\coth\theta\frac{\partial}{\partial\phi}-\operatorname{csch}\theta\frac{\partial}{\partial\psi}\right)^2\right]+\gamma\frac{\partial^2}{\partial\phi^2}-\beta\frac{\partial}{\partial\phi}. \tag{5.11}$$

Similarly, for $\bar V$ of (4.19) we obtain the same expression (5.11) but with ϕ and ψ interchanged.

We must also express the quantities of interest $|A(l,x)|^2$ and $|B(l,x)|^2$, where A and B are given by (3.9), (3.10), in terms of the parametrization (5.1). A straightforward computation yields

$$\frac{|A(l,x)|^2}{P_{\max}}=\frac{1+\cosh\theta_2}{1+\cosh\theta_1\cosh\theta_2+\sinh\theta_1\sinh\theta_2\cos(\phi_1+\psi_2)}, \tag{5.12}$$

$$\frac{|B(l,x)|^2}{P_{\max}}=\frac{-1+\cosh\theta_2}{1+\cosh\theta_1\cosh\theta_2+\sinh\theta_1\sinh\theta_2\cos(\phi_1+\psi_2)}, \tag{5.13}$$

$$P_{\max}=|E_g|^2\cosh^2\frac{\theta_g}{2}=\frac{|e_g|^2}{4\,\mathrm{Re}(Z_g)}. \tag{5.14}$$

Here we have omitted arguments on the random functions (θ_i,ϕ_i,ψ_i), $i=1,2$, to simplify the notation. The last equality in (5.14) follows from (2.8), (2.9), and (5.3). $P_{\max}$ is the maximum power available from the generator. The quantities on the left sides of (5.12) and (5.13) are therefore normalized incident and reflected powers, respectively, at position x, $0\le x\le l$, for a line of length l. Note that in both (5.12) and (5.13) ψ_1 and ϕ_2 are absent. This leads to substantial simplification in the analysis that follows.

Let $f(\theta,\phi)$ be a bounded continuous function on $SU(1,1)$ independent of ψ. Such a function may be regarded as a function on the hyperbolic disc which is identified with the space of right cosets[18] of $SU(1,1)$ modulo the subgroup generated by $e^{\psi\eta_1}$. The limit theorem of Sec. 4 in conjunction with (5.11) and the ψ independence of f imply that

$$\lim_{\epsilon\to 0}E\{f(\theta_1^{(\epsilon)}(\sigma),\phi_1^{(\epsilon)}(\sigma))\}=u_1(\sigma,\theta,\phi)\Big|_{\substack{\theta=\theta_g\\ \phi=\phi_g}}, \tag{5.15}$$

where u_1 satisfies the Cauchy problem

$$\frac{\partial u_1}{\partial\sigma}=Mu_1,\quad \sigma>0,\quad u_1(0,\theta,\phi)=f(\theta,\phi), \tag{5.16}$$

$$M=\Delta+L,\quad \Delta=\alpha\left(\frac{\partial^2}{\partial\theta^2}+\coth\theta\frac{\partial}{\partial\theta}+\operatorname{csch}^2\theta\frac{\partial^2}{\partial\phi^2}\right),$$
$$L=(\gamma+\alpha)\frac{\partial^2}{\partial\phi^2}-\beta\frac{\partial}{\partial\phi},\quad 0<\theta<\infty,\quad 0\le\phi<2\pi. \tag{5.17}$$

Note that Δ is the Laplace-Beltrami operator on the hyperbolic disc.[18] L is a simple differential operator with constant coefficients which commutes with Δ.

As indicated at the end of Sec. 4 the limit of the process $(\theta_2^{(\epsilon)},\phi_2^{(\epsilon)},\psi_2^{(\epsilon)})$ can be characterized in the same manner as (5.15). We have to replace M by the operator that corresponds to $\bar V$ of (4.19). Thus, we have the following result. Let $f(\theta,\psi)$ be a bounded continuous function on $SU(1,1)$ independent of ϕ. Then

$$\lim_{\epsilon\to 0}E\{f(\theta_2^{(\epsilon)}(\tau,\sigma),\psi_2^{(\epsilon)}(\tau,\sigma))\}=u_2(\tau-\sigma,\theta,\psi)\Big|_{\substack{\theta=\theta_l\\ \psi=\psi_l}}, \tag{5.18}$$

where $u_2(\sigma,\theta,\psi)$ satisfies the Cauchy problem (5.16), (5.17) with ϕ everywhere replaced by ψ.

Both u_1 and u_2 can be expressed conveniently in terms of the transition probability density $P(\sigma;\theta,\phi;\theta_0,\phi_0)$ of the process $(\theta^{(0)}(\sigma),\phi^{(0)}(\sigma))$ which at $\sigma=0$ takes the value (θ_0,ϕ_0). P is a density relative to the volume element $\sinh\theta\,d\theta\,d\phi$ and satisfies the forward equation

$$\frac{\partial P}{\partial\sigma}=(\Delta+L^*)P,\quad \sigma>0,$$
$$P(0;\theta,\phi;\theta_0,\phi_0)=\delta(\cosh\theta-\cosh\theta_0)\cdot\delta(\phi-\phi_0). \tag{5.19}$$

Here L^* is given by

$$L^*=(\gamma+\alpha)\frac{\partial^2}{\partial\phi^2}+\beta\frac{\partial}{\partial\phi}, \tag{5.20}$$

and δ denotes the ordinary delta function. The solution of (5.20) is

$$P(\sigma;\theta,\phi;\theta_0,\phi_0)=\sum_{m=-\infty}^{\infty}e^{-[m^2(\gamma+\alpha)+im\beta]\sigma}\left(\frac{e^{im(\phi-\phi_0)}}{2\pi}\right.$$
$$\times\int_0^\infty e^{-[\nu^2+(1/4)]\alpha\sigma}\frac{\nu\sinh\pi\nu}{\pi}\Gamma(\tfrac12-|m|+i\nu) \tag{5.21}$$
$$\left.\times\Gamma(\tfrac12-|m|-i\nu)P^{|m|}_{-1/2+i\nu}(\cosh\theta)P^{|m|}_{-1/2+i\nu}(\cosh\theta_0)d\nu\right).$$

Here $w(v)=P^m_{-1/2+i\nu}(v)$ is the conical Legendre function and it satisfies the equation[20]

$$\frac{d}{dv}\left((v^2-1)\frac{dw}{dv}\right)-\frac{m^2}{v^2-1}w=-(\nu^2+\tfrac14)w,\quad v>1. \tag{5.22}$$

We can now invoke the independence of the limits of Y_1 and Y_2 in (4.17), (4.18) and apply the above results to $|A|^2$ and $|B|^2$ of (5.12), (5.13). This yields the main result of this section

$$\frac{1}{P_{\max}}\lim_{\epsilon\to 0} E\left\{\begin{matrix}|A(\tau/\epsilon^2,\sigma/\epsilon^2)|^2\\ |B(\tau/\epsilon^2,\sigma/\epsilon^2)|^2\end{matrix}\right\} = \int_0^\infty\int_0^{2\pi}\int_0^\infty\int_0^{2\pi} \frac{\pm 1+\cosh\theta_2}{1+\cosh\theta_1\cosh\theta_2+\sinh\theta_1\sinh\theta_2\cos(\phi_1+\psi_2)}$$
$$\times P(\sigma;\theta_1,\phi_1;\theta_g,\phi_g)P(\tau-\sigma;\theta_2,\psi_2;\theta_l,\psi_l)\sinh\theta_1\sinh\theta_2 d\theta_1 d\phi_1 d\theta_2 d\psi_2. \quad (5.23)$$

Here the + sign corresponds to $|A|^2$ and the − sign to $|B|^2$.

When $x = l$ then $\sigma = \tau = \epsilon^2 l$ and $e^{i\phi^{(0)}}(\tau)\tanh\theta^{(0)}(\tau)/2$ is the right reflection coefficient[19] of the line. Thus, (5.21) is the joint transition probability density of the amplitude and phase of the reflection coefficient in the usual asymptotic limit. Note that in the matched case $\theta_0 = 0$, the phase is uniformly distributed and in any case when τ is large the phase is approximately uniformly distributed. To obtain the joint transition probability density of the left reflection coefficient $-e^{i\psi^{(0)}}(\tau)\tanh\theta^{(0)}(\tau)/2$, we note from (5.11) that we must solve (5.19) with L^* omitted and with ϕ and ϕ_0 replaced by ψ and ψ_0. The solution is

$$P(\tau;\theta,\psi;\theta_0,\psi_0) = \sum_{m=-\infty}^{\infty}\frac{e^{im(\psi-\psi_0)}}{2\pi}\cdot\int_0^\infty e^{-(\nu^2+1/4)\alpha\tau}\frac{\nu\sinh\pi\nu}{\pi}$$
$$\times\Gamma(\tfrac{1}{2}-|m|+i\nu)\Gamma(\tfrac{1}{2}-|m|-i\nu)$$
$$\times P^{|m|}_{-1/2+i\nu}(\cosh\theta)P^{|m|}_{-1/2+i\nu}(\cosh\theta_0)d\nu. \quad (5.24)$$

Again, if $\theta_0 = 0$, the matched case, the phase is uniformly distributed in our asymptotic limit for all $\tau \geq 0$.

However, contrary to what happens with the right reflection coefficient the phase is not uniform or approximately uniform for $\tau \geq 0$ in the mismatched case. Results about the phase of reflection coefficients have been reported recently in Refs. 21-23.

6. HEAT EQUATION FOR THE TOTAL POWER IN THE INTERIOR

According to the remarks at the end of Sec. 2, it is convenient to compute the expectations of $|A|^2 + |B|^2$ and $|A|^2 - |B|^2$ in the asymptotic limit of Sec. 4. In this section we consider the expectation of the total power $J(\tau,\xi)$, $\tau \geq 0$, $-\tau/2 \leq \xi \leq \tau/2$, defined by

$$J(\tau,\xi) = P_{\max}^{-1}\lim_{\epsilon\to 0} E\{|A(\tau/\epsilon^2,(\xi+\tau/2)/\epsilon^2)|^2 + |B(\tau/\epsilon^2,(\xi+\tau/2)/\epsilon^2)|^2\}. \quad (6.1)$$

Clearly $J = J(\tau,\xi;\theta_g,\phi_g;\theta_l,\psi_l)$ but we shall not indicate the dependence on the generator and load parameters explicitly. The variable $\tau = \epsilon^2 l$ is the scaled line length, as in the two previous sections, and ξ is scaled distance from the midpoint $\tau/2$ of the line.

The purpose of this section is to prove the following theorem which generalizes an observation of Gazaryan.[3]

Theorem: Let $J(\tau,\xi)$ be defined by (6.1) and α,β,γ by (4.12). Then $J(\tau,\xi)$ is defined for $\tau \geq 0$ and $-\infty < \xi < \infty$. Moreover,

$$J(\tau,\xi) = \sum_{m=-\infty}^{\infty} J_m(\tau,\xi), \quad (6.2)$$

where

$$\frac{\partial}{\partial\tau}J_m(\tau,\xi) = \frac{1}{4\alpha}\frac{\partial^2}{\partial\xi^2}J_m(\tau,\xi) - \kappa_m J_m(\tau,\xi), \quad (6.3)$$
$$\tau > 0, \quad -\infty < \xi < \infty,$$

and

$$J_m(0,\xi) = \frac{\pi}{2i}e^{im(\phi_g+\psi_l)}e^{-\alpha\xi}\int_{-\infty}^{\infty}\frac{\sinh\pi t\Gamma(\frac{1}{2}-|m|+it)}{\cosh^2\pi t\Gamma(\frac{1}{2}+|m|+it)}$$
$$\times P^{|m|}_{-1/2+it}(\cosh\theta_g)\{(\tfrac{1}{2}+it-|m|)P^{|m|}_{-3/2-it}(\cosh\theta_l)$$
$$\times e^{-i2\alpha\xi t} - (\tfrac{1}{2}-it-|m|)P^{|m|}_{-3/2+it}(\cosh\theta_l)e^{i2\alpha\xi t}\}dt, \quad -\infty < \xi < \infty, \quad (6.4)$$

$$\kappa_m = (\gamma+\alpha)m^2 + i\beta m. \quad (6.5)$$

Proof: Let $u = \cosh\theta_1$, $v = \cosh\theta_2$, $\phi = \phi_1$, $\psi = \psi_2$ and define $f(u,v,\phi,\psi)$ by

$$f(u,v,\phi,\psi) = 2v/[1+uv+\sqrt{u^2-1}\sqrt{v^2-1}\cos(\phi+\psi)]. \quad (6.6)$$

Then, from (5.23) and (6.1) it follows that

$$J(\tau,\xi) = \int_1^\infty\int_0^{2\pi}\int_1^\infty\int_0^{2\pi} f(u,v,\phi,\psi)P(\tau/2+\xi;u,\phi;u_g,\phi_g)$$
$$\times P(\tau/2-\xi;v,\psi;v_l,\psi_l)dud\phi dvd\psi, \quad (6.7)$$

where $u_g = \cosh\theta_g$, $v_l = \cosh\theta_l$.

We observe now that the denominator on the right side of (6.6) is related to the law of cosines on the hyperbolic disc. Then we define $\zeta \geq 0$ by

$$\cosh\zeta = uv - \sqrt{u^2-1}\sqrt{v^2-1}\cos(\phi+\psi+\pi), \quad (6.8)$$

and rewrite (6.6) in the form

$$f(u,v,\phi,\psi) = 2v/(1+\cosh\zeta). \quad (6.9)$$

We proceed next to obtain a Fourier expansion of f. For this we need the following facts concerning Legendre functions:[20]

$$\frac{2}{1+\cosh\zeta} = \pi\int_{-\infty}^{\infty}\frac{t\sinh\pi t}{\cosh^2\pi t}P_{-1/2+it}(\cosh\zeta)dt, \quad (6.10)$$

$$P_\nu(\cosh\zeta) = \sum_{m=-\infty}^{\infty}\frac{\Gamma(\nu-|m|+1)}{\Gamma(\nu+|m|+1)}P_\nu^{|m|}(u)P_\nu^{|m|}(v)e^{im(\phi+\psi)}, \quad (6.11)$$

$$P_\nu^m(u) = P_{-\nu-1}^m(u), \quad (6.12)$$

$$uP_\nu^{|m|}(u) = \frac{1}{2\nu+1}[(\nu-|m|+1)P_{\nu+1}^{|m|}(u) + (\nu+|m|)P_{\nu-1}^{|m|}(u)]. \quad (6.13)$$

On using (6.10)-(6.13) in (6.9) we find, after some rearrangements, that

$$\frac{2v}{1+\cosh\zeta} = \sum_{m=-\infty}^{\infty}Q_{|m|}(u,v)e^{im(\phi+\psi)}, \quad (6.14)$$

where

$$Q_{|m|}(u,v)=\frac{\pi}{2i}\int_{-\infty}^{\infty}\frac{\sinh\pi t\Gamma(\frac{1}{2}-|m|+it)}{\cosh^2\pi t\Gamma(\frac{1}{2}+|m|+it)}P^{|m|}_{-1/2+it}(u)\times\{[-(|m|-\tfrac{1}{2})+it]P^{|m|}_{-3/2-it}(v)+[(|m|-\tfrac{1}{2})+it]P^{|m|}_{-3/2+it}(v)\}dt. \quad (6.15)$$

We return now to (6.7), substitute for f the expression (6.14) and perform the ϕ and ψ integrations. This yields

$$J(\tau,\xi)=\sum_{m=-\infty}^{\infty}J_m(\tau,\xi), \quad (6.16)$$

where

$$J_m(\tau,\xi)=\int_1^\infty\int_1^\infty Q_{|m|}(u,v)\hat{P}_m(\tau/2+\xi)\hat{P}_m(\tau/2-\xi)dudv \quad (6.17)$$

and $\hat{P}_m(\sigma)=\hat{P}_m(\sigma;u;u_0,\phi_0)$ is defined by

$$\hat{P}_m(\sigma)=\int_0^{2\pi}P(\sigma;u,\phi;u_0,\phi_0)e^{im\phi}d\phi. \quad (6.18)$$

$P(\sigma)$ satisfies (5.19) and it is given explicitly by (5.21). From (5.19) and (6.18) it follows that $\hat{P}_m(\sigma;u;u_0,\phi_0)$ satisfies the Cauchy problem

$$\frac{\partial}{\partial\sigma}\hat{P}_m=\alpha\Delta_{1,m}\hat{P}_m-\kappa_m\hat{P}_m,\qquad \hat{P}_m(0)=e^{im\phi_0}\delta(u-u_0), \quad (6.19)$$

$$\Delta_{1,m}=\frac{\partial}{\partial u}\left[(u^2-1)\frac{\partial}{\partial u}\right]-\frac{m^2}{u^2-1},\qquad u>1, \quad (6.20)$$

and κ_m is given by (6.5).

Let us rewrite (6.17) using operator notation, (6.19) and (6.20). We have

$$J_m(\tau,\xi)=e^{im(\phi_g+\psi_l)}\cdot\{e^{(\tau/2+\xi)(\alpha\Delta_{1,m}-\kappa_m)+(\tau/2-\xi)(\alpha\Delta_{2,m}-\kappa_m)}Q_{|m|}\}(u_g,v_l)$$
$$=e^{im(\phi_g+\psi_l)-\tau\kappa_m}\{e^{\alpha(\tau/2+\xi)\Delta_{1,m}+\alpha(\tau/2-\xi)\Delta_{2,m}}Q_{|m|}\}(u_g,v_l). \quad (6.21)$$

In (6.21) $\Delta_{1,m}$ acts only on the first argument of $Q_{|m|}(u,v)$ and $\Delta_{2,m}$ is the same as $\Delta_{1,m}$ but acts on the second argument of $Q_{|m|}$. Thus $\Delta_{1,m}$ and $\Delta_{2,m}$ commute. By direct differentiation of (6.21) it follows that (6.3) holds provided that

$$(\Delta_{1,m}+\Delta_{2,m})Q_{|m|}=\tfrac{1}{2}(\Delta_{1,m}-\Delta_{2,m})^2Q_{|m|}. \quad (6.22)$$

To check that (6.22) is indeed true we need only employ the following relations in the definition (6.15) of $Q_{|m|}$:

$$\Delta_{1,m}P^{|m|}_{-1/2+it}(u)P^{|m|}_{-3/2\pm it}(v)=-(t^2+\tfrac{1}{4})P^{|m|}_{-1/2+it}(u)P^{|m|}_{-3/2\pm it}(v), \quad (6.23)$$

$$\Delta_{2,m}P^{|m|}_{-1/2+it}(u)P^{|m|}_{-3/2\pm it}(v)=-(t^2\pm 2it-\tfrac{3}{4})P^{|m|}_{-1/2+it}(u)P^{|m|}_{-3/2\pm it}(v). \quad (6.24)$$

This completes the proof of (6.3).

To compute the initial value $J_m(0,\xi)$ we set $\tau=0$ in (6.21) and obtain

$$J_m(0,\xi)=e^{im(\phi_g+\psi_l)}\{e^{\alpha\xi(\Delta_{1,m}-\Delta_{2,m})}Q_{|m|}\}(u_g,v_l). \quad (6.25)$$

Using (6.23) and (6.24) in the definition (6.15) of $Q_{|m|}$, it follows that

$$([\Delta_{1,m}-\Delta_{2,m}]Q_{|m|})(u_g,v_l)=\frac{\pi}{2i}\int_{-\infty}^{\infty}\frac{\sinh\pi t\Gamma(\frac{1}{2}-|m|+it)}{\cosh^2\pi t\Gamma(\frac{1}{2}+|m|+it)}P^{|m|}_{-1/2+it}(u_g)\times\{[-(|m|-\tfrac{1}{2})+it]P^{|m|}_{-3/2-it}(v_l)(-1-2it)+[(|m|-\tfrac{1}{2})+it]P^{|m|}_{-3/2+it}(v_l)(-1+2it)\}dt. \quad (6.26)$$

By formally expanding the exponential in (6.25) and using (6.26) and its iterates, and performing a few rearrangements, the result (6.4) follows. This completes the proof of the theorem.

When the line is matched then $\Gamma_g=\Gamma_l=0$ and hence, from (5.3), $\theta_g=\theta_l=0$. Since $P^{|m|}_\nu(1)=0$, $m\neq 0$ it follows from (6.5) that

$$J_m(\tau,\xi)=0,\qquad m\neq 0. \quad (6.27)$$

Thus $J(\tau,\xi)=J_0(\tau,\xi)$ and

$$\frac{\partial J}{\partial\tau}=\frac{1}{4\alpha}\frac{\partial^2 J}{\partial\xi^2},\qquad \tau>0, \quad (6.28)$$

$$J(0,\xi)=\pi e^{-\alpha\xi}\int_{-\infty}^{\infty}\frac{t\sinh\pi t}{\cosh^2\pi t}\left(\cos 2\alpha\xi t-\frac{\sin 2\alpha\xi t}{2t}\right)dt$$
$$=1-\tanh\alpha\xi-\frac{\alpha\xi}{\cosh^2\alpha\xi}. \quad (6.29)$$

This is the result Gazaryan[3] obtained for a very special kind of fluctuation process $\mu(x)$, and so the theorem above is indeed a generalization of it.

Equations (6.3) can be solved explicitly in an elementary manner. Thus, we have

$$J(\tau,\xi)=\sum_{m=-\infty}^{\infty}e^{im(\phi_g+\psi_l)-\kappa_m\tau}\sqrt{\frac{\alpha}{\pi\tau}}\int_{-\infty}^{\infty}e^{-\alpha(\xi-\eta)^2/\tau}\tilde{J}_m(0,\eta)d\eta. \quad (6.30)$$

Here $\tilde{J}_m(0,\xi)$ is identical with (6.4) with the factor $e^{im(\phi_g+\psi_l)}$ omitted. Using (6.4) in (6.30), performing the η integration and rearranging yields

$$J(\tau,\xi)=e^{\alpha(\tau/2)-\alpha\xi}\sum_{m=-\infty}^{\infty}e^{im(\phi_g+\psi_l)-\kappa_m\tau}\times\pi\int_{-\infty}^{\infty}e^{-\alpha\tau(t^2+\frac{1}{4})}\frac{t\sinh\pi t\Gamma(\frac{1}{2}-|m|+it)}{\cosh^2\pi t\Gamma(\frac{1}{2}+|m|+it)}\times P^{|m|}_{-1/2+it}(\cosh\theta_g)\{\cosh\theta_l\cos\alpha t(\tau-2\xi)P^{|m|}_{-1/2+it}(\cosh\theta_l)-\frac{\sin\alpha t(\tau-2\xi)}{2t}[(\tfrac{1}{2}-|m|+it)P^{|m|}_{-3/2-it}(\cosh\theta_l)+(\tfrac{1}{2}-|m|-it)P^{|m|}_{-3/2+it}(\cosh\theta_l)]\}dt. \quad (6.31)$$

In Sec. 5, below (5.5) we observed that ψ_l depends explicitly on l. Since $l=\tau/\epsilon^2$ is going to infinity as $\epsilon\to 0$ only $J_0(\tau,\xi)$ in (6.2) is meaningful physically. The rapid phase oscillations due to ψ_l will average to zero, within the range of parameters considered here, in any measuring process. From (6.31) we obtain the following representation for $J_0(\tau,\xi)$:

$$J_0(\tau,\xi)=e^{\alpha(\tau/2)-\alpha\xi}\pi\int_{-\infty}^{\infty}e^{-\alpha\tau(t^2+1/4)}\frac{t\sinh\pi t}{\cosh^2\pi t}P_{-1/2+it}(\cosh\theta_g)\times\left\{\cosh\theta_l\cos\alpha t(\tau-2\xi)P_{-1/2+it}(\cosh\theta_l)-\frac{\sin\alpha t(\tau-2\xi)}{2t}[(\tfrac{1}{2}+it)P_{-3/2-it}(\cosh\theta_l)+(\tfrac{1}{2}-it)P_{-3/2+it}(\cosh\theta_l]\right\}dt. \quad (6.32)$$

In Figs. 2-4 we plot $J_0(\tau, \xi)$ as a function of ξ, $-\tau/2 \le \xi \le \tau/2$ for various values of $\tau \ge 0$, $\theta_g \ge 0$, $\theta_l \ge 0$. The graphs were obtained by evaluating (6.32) numerically.

The above theorem is somewhat surprising and one is led to inquire if it could have been anticipated without actually performing the computations. This is not an easy task however, because a great deal of simplification and decoupling occurs in the asymptotic limit. Furthermore, within the context of the asymptotic limit, the result appears as a somewhat peculiar property of conditional expectations of certain functionals of, essentially, Brownian motion on the hyperbolic disc.

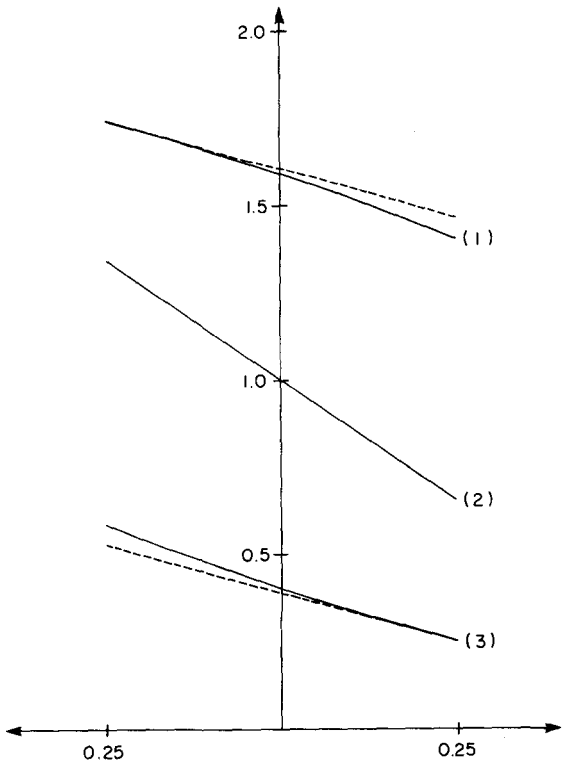


FIG. 2. Here we plot J_0 (solid lines) and J_s (broken lines) versus $\alpha\xi$, $-\alpha\tau/2 \le \alpha\xi \le \alpha\tau/2$, for $\alpha\tau = 0.5$. The curves labeled by (1), (2), (3) correspond to $\theta_g = 0$ and $\theta_l = 2.4$, $\theta_g = \theta_l = 0$, $\theta_g = 2.4$ and $\theta_l = 0$, respectively. In this figure the solid and broken lines coalesce when $\theta_g = \theta_l = 0$.

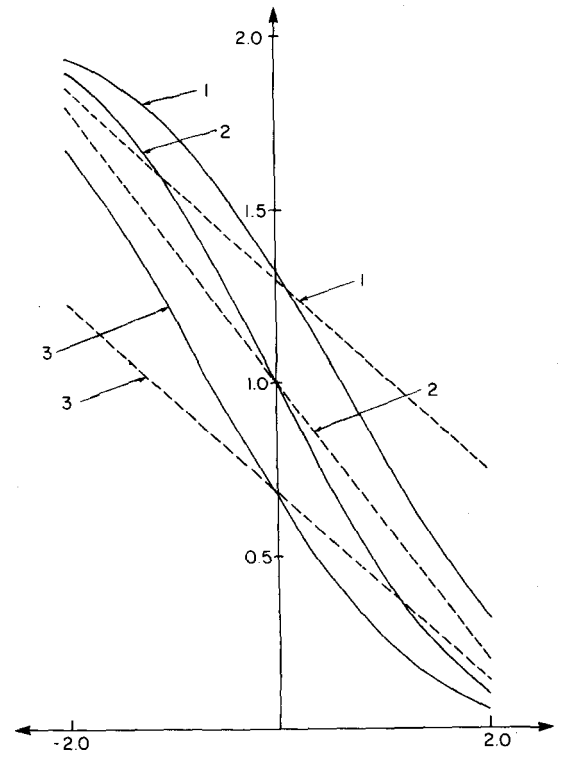


FIG. 3. For $\alpha\tau = 4$ (see Fig. 2).

7. TRANSMISSION COEFFICIENTS

We shall compute here the expectation of $P_{\max}^{-1}(|A|^2 - |B|^2)$ in the asymptotic limit of Sec. 4. It can be verified that this quantity is independent of x, $0 \le x \le l$. As noted in (2.18) it represents the power flux through the line and we call it the power transmission coefficient. Thus we set

$$PT(\tau) = P_{\max}^{-1} \lim_{\epsilon \to 0} E\{|A(\tau/\epsilon^2, \tau/\epsilon^2)|^2 - |B(\tau/\epsilon^2, \tau/\epsilon^2)|^2\}, \tag{7.1}$$

where we have chosen to let $x = l$. From (5.23), (5.19), and (7.1), we obtain

$$PT(\tau) = \int_1^\infty \int_0^{2\pi} g(u, v_l, \phi; \psi_l) P(\tau; u, \phi; u_g, \phi_g) du d\phi, \quad \tau \ge 0. \tag{7.2}$$

Here P is given by (5.21), $u = \cosh\theta$, $u_g = \cosh\theta_g$, $v_l = \cosh\theta_l$ and

$$g(u, v, \phi, \psi) = 2[1 + uv + \sqrt{u^2 - 1}\sqrt{v^2 - 1}\cos(\phi + \psi)]^{-1}. \tag{7.3}$$

It is not necessary to perform any computations in evaluating (7.2) for $PT(\tau)$ because we can use $J(\tau, \xi)$ of the previous section as follows From (2.7) we notice that at $x = l$; we have

$$|A|^2 + |B|^2 = \left(\frac{1 + |\Gamma_l|^2}{1 - |\Gamma_l|^2}\right)(|A|^2 - |B|^2). \tag{7.4}$$

Thus, from (5.3), (6.1), and (7.1)

$$PT(\tau) = \operatorname{sech}\theta_l J(\tau, \tau/2). \tag{7.5}$$

Setting $\xi = \alpha/2$ in (6.31) yields the desired result:

$$PT(\tau) = \sum_{m=-\infty}^{\infty} e^{im(\phi_g + \psi_l) - \kappa_m \tau \pi} \int_{-\infty}^{\infty} e^{-\alpha\tau(t^2 + 1/4)} \frac{t \sinh \pi t}{\cosh^2 \pi t} \times \frac{\Gamma(\frac{1}{2} - |m| + it)}{\Gamma(\frac{1}{2} + |m| + it)} \cdot P_{-1/2+it}^{|m|}(\cosh\theta_g) \times P_{-1/2+it}^{|m|}(\cosh\theta_l) dt. \tag{7.6}$$

Here κ_m and α are defined by (6.5) and (4.12), respectively. Note that (7.14) is symmetric in the load and generator parameters:

$$PT(\tau; \theta_g, \phi_g; \theta_l, \psi_l) = PT(\tau; \theta_l, \psi_l; \theta_g, \phi_g). \tag{7.7}$$

In order to compare (7.6) with the results of J. A. Morrison[15] we must identify the load and generator parameters with those of problem (2.14). This amounts to expressing (2.16) and (2.17) in polar coordinates (5.3) and using (5.14). A simple calculation yields

$$\cosh\theta_g = \tfrac{1}{2}(n_1 + 1/n_1), \quad \phi_g = \pi H(n_1 - 1), \tag{7.8}$$

$$\cosh\theta_l = \tfrac{1}{2}(n_2 + 1/n_2), \quad \psi_l = 2ikl + \pi H(n_2 - 1) + \pi, \tag{7.9}$$

$$P_{\max} = n_1. \tag{7.10}$$

Here $H(x)$ denotes the Heaviside unit step function. Upon using (7.8)-(7.10) in (7.6) we recover Morrison's formula[15] (4.17) by employing the identity[20]

$$\Gamma(\tfrac{1}{2} - |m| - it)\Gamma(\tfrac{1}{2} + |m| + it) = (-1)^m \pi/\cosh\pi t, \tag{7.11}$$

and noting that his β_0 corresponds to $-k$ here.

When the line is matched to the generator and the load then $\theta_g = \theta_l = 0$ and (7.6) simplifies to

$$PT\Big|_{\substack{\theta_g=0\\ \theta_l=0}} = 2\pi \int_0^\infty e^{-\alpha\tau(t^2+1/4)} \frac{t\sinh\pi t}{\cosh^2\pi t}\,dt, \quad \tau \geq 0. \tag{7.12}$$

Using (7.5), (6.28), and (6.29) we find another representation of the result (7.12):

$$\begin{aligned} PT(\tau)\Big|_{\substack{\theta_g=0\\ \theta_l=0}} &= \sqrt{\alpha/\pi\tau}\int_{-\infty}^{\infty} e^{-\alpha\tau(\tau/2-\eta)^2} \\ &\quad \times (1 - \tanh\alpha\eta - \alpha\eta/\cosh^2\alpha\eta)d\eta \\ &= \frac{4e^{-\alpha\tau/4}}{\sqrt{\pi}}\int_0^\infty \frac{\eta^2 e^{-\eta^2}d\eta}{\cosh(\eta\sqrt{\alpha\tau})}. \end{aligned} \tag{7.13}$$

Formulas (7.12) and (7.13) have been obtained previously by a variety of methods.[24,25,26]

8. COMPARISON WITH RADIATIVE TRANSPORT THEORY

Radiative transport theory is a phenomenological theory that considers the transport of radiation from one region of an inhomogeneous medium to another as an incoherent scattering phenomenon disregarding the wave nature of the transfer process. We shall describe briefly this theory in connection with problem (2.10)-(2.13) first when $n_1 = n_2 = 1$ and then in the general case.

Let $I^+(t,\tau,\sigma)$, $t \geq 0$, $0 \leq \sigma \leq \tau$, represent the intensity of radiation at time t and location σ propagating in the positive σ direction through an inhomogeneous medium occupying the interval $[0,\tau]$. Let $I^-(t,\tau,\sigma)$ represent the intensity propagating in the negative σ direction. Elementary physical arguments[5] lead to the following conservation equations for $I^\pm$:

$$\begin{aligned} &\frac{1}{v}\frac{\partial I^+}{\partial t} + \frac{\partial I^+}{\partial\sigma} = -\alpha I^+ + \beta I^-, \\ &\frac{1}{v}\frac{\partial I^-}{\partial t} - \frac{\partial I^-}{\partial\sigma} = -\alpha I^- + \beta I^+, \quad 0 \leq \sigma \leq \tau, \quad t \geq 0, \end{aligned} \tag{8.1}$$

$$I^\pm(0,\tau,\sigma) = I_0^\pm(\sigma), \tag{8.2}$$

$$I^+(t,\tau,0) = 1, \quad I^-(t,\tau,\tau) = 0. \tag{8.3}$$

Here v denotes the transport velocity, α and β are transport coefficients characteristic of the scattering, absorptive and emittive properties of the medium and (8.3) has been chosen so that radiation of unit intensity is incident on the medium from the left. In the steady state regime and for a conservative medium we set the time derivatives equal to zero in (8.1) and $\alpha = \beta$. Thus $I^\pm(\tau,\sigma)$ satisfy the equations

$$\begin{aligned} &\frac{d}{d\sigma}I^+ = -\alpha(I^+ - I^-), \\ &-\frac{d}{d\sigma}I^- = \alpha(I^+ - I^-), \quad 0 \leq \sigma \leq \tau, \end{aligned} \tag{8.4}$$

$$I^+(\tau,0) = 1, \quad I^-(\tau,\tau) = 0. \tag{8.5}$$

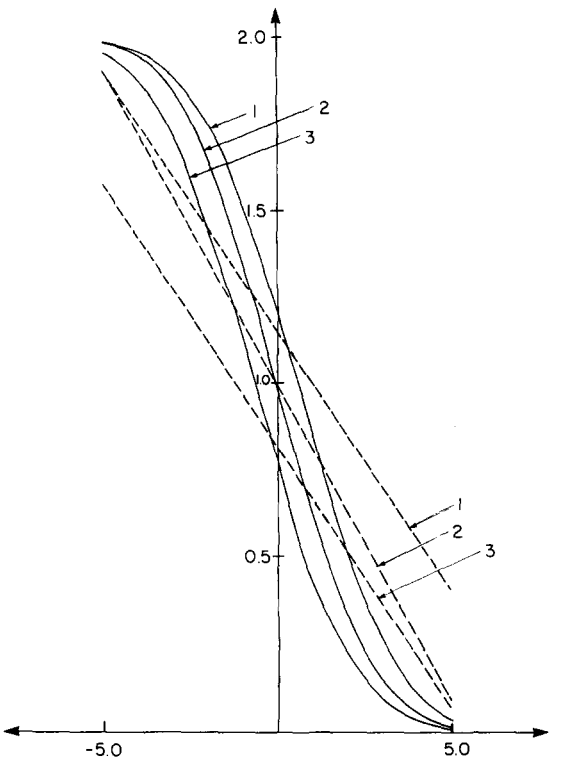


FIG. 4. For $\alpha\tau = 10$ (see Fig. 2).

The quantity $I^+ + I^-$ is the total radiation and $I^+ - I^-$ is the flux of radiation. The latter depends on τ only. Equations (8.4), (8.5) are elementary and their solution is

$$I^+(\tau,\sigma) = \frac{1+\alpha(\tau-\sigma)}{1+\alpha\tau}, \quad I^-(\tau,\sigma) = \frac{\alpha(\tau-\sigma)}{1+\alpha\tau}, \quad 0 \leq \sigma \leq \tau. \tag{8.6}$$

The above theory is entirely phenomenological and it was first employed by Schuster;[5] see also Ref. 10, 27. A general treatment of radiative transport theory can be found in Ref. 6.

The question that concerns us here is the following. What is the relation, if any, between (8.4), (8.5) and the stochastic boundary value problem (2.6), (2.7)? We assume here that $E_g = 1$, $\Gamma_g = \Gamma_l = 0$ in (2.7). Several investigators[7,8,9] have given heuristic arguments indicating that $I^+(\tau,\sigma)$ and $I^-(\tau,\sigma)$ should be $\lim_{\epsilon\to 0} E\{|A(\tau/\epsilon^2, \sigma/\epsilon^2)|^2\}$ and $\lim_{\epsilon\to 0} E\{|B(\tau/\epsilon^2,\sigma/\epsilon^2)|^2\}$, respectively, under more or less the same conditions as stated in Sec. 4 and with α in (8.4) given by (4.12). This would be a very satisfactory answer to our question, if it were correct, because (8.4) and (8.5) are very simple equations. Unfortunately, it cannot be correct without further restrictions since it does not agree with the results of Secs. 6 and 7 which follow from a rigorous mathematical theory.

In order to compare transport theory to the general mismatched case we must change the boundary condition in (8.5). If we accept the correspondence between A, B, and $I^\pm$ discussed above then, in view of (2.7), (8.5) should be

$$\begin{aligned} I^+(\tau,0) &= |E_g|^2 + |\Gamma_g|^2 I^-(\tau,0), \\ I^-(\tau,\tau) &= |\Gamma_l|^2 I^+(\tau,\tau). \end{aligned} \tag{8.7}$$

Solving (8.4) and (8.7), we obtain

$$J_s(\tau,\xi) = P_{\max}^{-1}[I^+(\tau,\tau/2+\xi) + I^-(\tau,\tau/2+\xi)]$$
$$= 1 - \left(\frac{\frac{1}{2}(\cosh\theta_g - \cosh\theta_l) + 2\alpha\xi}{\frac{1}{2}(\cosh\theta_g + \cosh\theta_l) + \alpha\tau}\right),$$
$$-\tau/2 \le \xi \le \tau/2, \tag{8.8}$$

$$PT_s(\tau) = P_{\max}^{-1}[I^+(\tau,\tau/2+\xi) - I^-(\tau,\tau/2+\xi)]$$
$$= \frac{1}{\frac{1}{2}(\cosh\theta_g + \cosh\theta_l) + \alpha\tau}. \tag{8.9}$$

Here we have used the subscript s (Schuster) to denote quantities obtained from transport theory, $P_{\max}$ is given by (5.14) and we have employed polar coordinates as in (5.3). When $\cosh\theta_g = \cosh\theta_l = 1$ the results (8.8), (8.9) reduce to the matched case (8.6).

Let us now compare (8.8) and (8.9) with the results obtained by the stochastic theory. Specifically, we compare $J_s(\tau,\xi)$ with $J_0(\tau,\xi)$ given by (6.32) and (8.9) with $PT_0(\tau)$ which is the $m = 0$ term in (7.6):

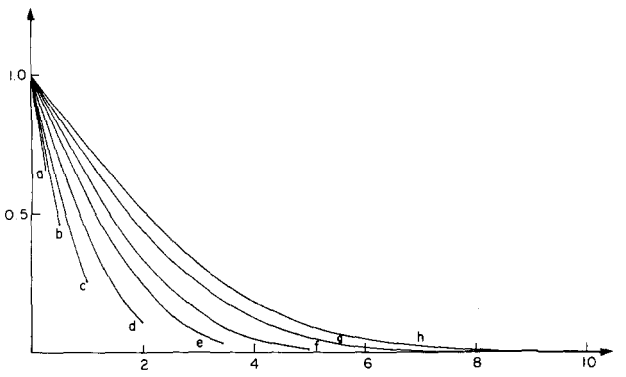


FIG. 5. Here we plot J_0 versus $\alpha\xi$, $0 \le \alpha\xi \le \alpha\tau/2$ for $\theta_g = \theta_l = 0$ and $\alpha\tau = .5$ (a), $\alpha\tau = 1$ (b), $\alpha\tau = 2$ (c), $\alpha\tau = 4$ (d), $\alpha\tau = 7$ (e), $\alpha\tau = 10$ (f), $\alpha\tau = 15$ (g), $\alpha\tau = 20$ (h).

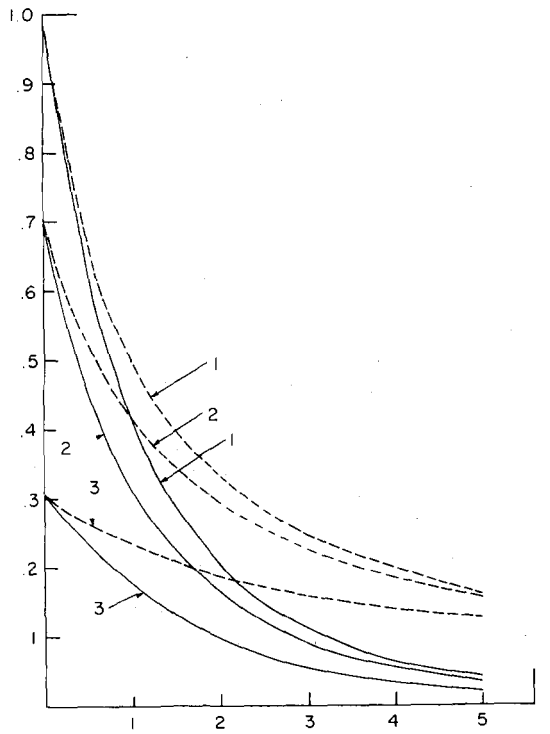


FIG. 6. Here we plot PT_0 (solid lines) and PT_s (broken lines) versus $\alpha\tau$, $0 \le \alpha\tau \le 5$. The curves labeled (1), (2), (3) correspond to $\theta_g = \theta_l = 0$, $\theta_g = 1.2$ and $\theta_l = 0$, $\theta_g = 2.4$ and $\theta_l = 0$, respectively.

$$PT_0(\tau) = \pi\int_{-\infty}^{\infty} e^{-\alpha\tau(t^2+1/4)}\frac{t\sinh\pi t}{\cosh^2\pi t}P_{-1/2+it}(\cosh\theta_g)$$
$$P_{-1/2+it}(\cosh\theta_l)dt. \tag{8.10}$$

The comparison of corresponding formulas is best done by examining the graphs shown in Figs. 2-6. However, for small $\alpha\tau$ we can expand both $PT_s(\tau)$ and $PT_0(\tau)$ in powers of $\alpha\tau$ and compare the first few terms:

$$PT_s(\tau) = \frac{2}{\cosh\theta_g + \cosh\theta_l} - \frac{4\alpha\tau}{(\cosh\theta_g + \cosh\theta_l)^2}$$
$$+ \frac{8(\alpha\tau)^2}{(\cosh\theta_g + \cosh\theta_l)^3} + \cdots \tag{8.11}$$

$$PT_0(\tau) = \frac{2}{\cosh\theta_g + \cosh\theta_l} - \frac{4(1+\cosh\theta_g\cosh\theta_l)\alpha\tau}{(\cosh\theta_g + \cosh\theta_l)^3}$$
$$+ \{4(\alpha\tau)^2[6 - (3 + \cosh\theta_l\cosh\theta_g)(\cosh^2\theta_g + \cosh^2\theta_l)$$
$$+ 2\cosh\theta_l\cosh\theta_g(3 + 2\cosh\theta_l\cosh\theta_g)]\}$$
$$\times(\cosh\theta_g + \cosh\theta_l)^{-5} + \cdots. \tag{8.12}$$

From (8.11) and (8.12) it follows that in the general case PT_s and PT_0 disagree even to first order in $\alpha\tau$. However, in the matched case, $\cosh\theta_g = \cosh\theta_l = 1$, PT_0 and PT_s agree to order $(\alpha\tau)^2$, but disagree in order $(\alpha\tau)^3$. From the figures we see that transport theory agrees fairly well with the stochastic theory in the matched case when $\alpha\tau$ is small. When $\alpha\tau$ is large or when there is impedance mismatch then we have fairly significant discrepancies.

It appears that in the asymptotic limit of Sec. 4 some phase information remains in $J_0(\tau,\xi)$ and $PT_0(\tau)$ and this leads to quantitative disagreement with the purely incoherent transport theory. On the other hand transport theory is much simpler. We may conclude however, that transport theory cannot be related systematically to stochastic wave equations in the simple manner suggested so far. This conclusion is supported by a recent analysis,[28] comparing the predictions of the two theories for power transmission in the matched case with computer-simulated results.

9. PROPAGATION OF PULSES

Up to now we have concerned ourselves exclusively with time harmonic dependence of the fields. The methods employed in a previous section can be used, however, to analyze problems with more general time dependence. In this section we examine the pulse problem in general and compute in particular the spreading of a Gaussian modulated pulse due to the fluctuations.

Let $\mathcal{E}_g(t)$ represent the generator voltage as a function of time and denote by $\mathfrak{A}(l,x;t)$ and $\mathfrak{B}(l,x,t)$ the time dependent incident and reflected waves at location x, $0 \le x \le l$, for a line of length $l \ge 0$ and at time $t \ge 0$. We introduce Fourier transforms as follows

$$e_g(\omega) = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty} e^{i\omega t}\mathcal{E}_g(t)dt, \tag{9.1}$$

$$e^{i\omega x/c}A(l,x;\omega) = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty} e^{i\omega t}\mathfrak{A}(l,x;t)dt \tag{9.2}$$

$$e^{-i\omega x/c}B(l,x;\omega) = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty} e^{i\omega t}\mathfrak{B}(l,x;t)dt. \tag{9.3}$$

We shall assume that $A(l,x;\omega)$ and $B(l,x;\omega)$ satisfy the boundary value problem (2.6), (2.7) where $k = \omega/c$ as in (2.4) and $e_g(\omega)$ in (2.8) is given by (9.1). Note also that, in general, $Z_g = Z_g(\omega)$ and $Z_l = Z_l(\omega)$ so that Γ_g and Γ_l in (2.9) are functions of ω. Similarly, the statistical characteristics of $\mu(x)$ in (2.8) may depend on ω so we will write $\mu(x;\omega)$ or $\mu(x;k)$.

The time dependent amplitudes $\mathfrak{A}(l,x;t)$ and $\mathfrak{B}(l,x;t)$ are obtained by using the inverse Fourier transform

$$\mathfrak{A}(l,x;t) = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty} e^{-i\omega(t-x/c)}A(l,x;\omega)d\omega, \tag{9.4}$$

$$\mathfrak{B}(l,x;t) = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty} e^{-i\omega(t+x/c)}B(l,x;\omega)d\omega. \tag{9.5}$$

The quantities of interest to us are the expected values of the time dependent incident and reflected powers. Before introducing these quantities, however, it is convenient to change our notation and express all quantities and integration variables in terms of $k = \omega/c$ rather than ω [cf. (2.14)]. Thus, we have the representations

$$E\{|\mathfrak{A}(l,x;t)|^2\} = \frac{c^2}{2\pi}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} e^{-i(k-k')(ct-x)}E\{A(l,x;k)\overline{A}(l,x;k')\}dkdk' \tag{9.6}$$

$$E\{|\mathfrak{B}(l,x;t)|^2\} = \frac{c^2}{2\pi}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} e^{-i(k-k')(ct+x)}E\{B(l,x;k)\overline{B}(l,x;k')\}dkdk' \tag{9.7}$$

From (9.6) and (9.7) it is apparent that the analysis of the pulse problem rests on knowledge of the joint statistics of the solution $A(l,x;k)$ and $B(l,x;k)$ of (2.6), (2.7) at two wave numbers k and k'.

In order to employ the limit theorem of section 4 for the joint statistics at two wave numbers we shall assume that the following conditions hold:

(i) $$P = \int_{-\infty}^{\infty} P_{\max}(k)dk < \infty, \qquad P_{\max}(k) = \frac{|e_g(k)|^2}{4\,\mathrm{Re}(Z_g(k))}, \tag{9.8}$$

(ii) $$P_{\max}(k) = 0, \qquad |k| > \Omega. \tag{9.9}$$

The first assumption asserts that the total power of the pulse is finite and the second asserts that the pulse is bandlimited. Let $\zeta > 0$ be fixed and define $\mathfrak{D}_\zeta$ and $\overline{\mathfrak{D}}_\zeta$ by

$$\mathfrak{D}_\zeta = \{(k,k') \,\big|\, |k-k'| \le \zeta\} \cup \{(k,k') \,\big|\, |k+k'| \le \zeta\},$$
$$\overline{\mathfrak{D}}_\zeta = \text{complement of } \mathfrak{D}_\zeta. \tag{9.10}$$

From the definition (5.14) of $P_{\max}(k)$ it follows that

$$|A(l,x;k)| \le P_{\max}^{1/2}(k), \qquad |B(l,x;k)| \le P_{\max}^{1/2}(k), \tag{9.11}$$

since $|A|^2$ and $|B|^2$ are the incident and reflected powers, respectively. From (9.8), (9.9), and (9.11) we conclude that we may replace the domain of integration in (9.6) and (9.7) by $\overline{\mathfrak{D}}_\zeta$ with an error which does not exceed $(4c^2P/\pi)\zeta$. Thus, for the purposes of the limit theorem, we shall assume that k and k' are distinct independently of $\epsilon \ge 0$ and, from (9.9), k and k' are bounded in absolute value.

We turn next to the computation of $E\{A(l,x;k)\overline{A}(l,x,k')\}$ and $E\{B(l,x;k)\overline{B}(l,x;k')\}$ in the asymptotic limit of Sec. 4. From (3.9) and (3.10) it follows that we must find the joint statistics of $Y_1(x,0;k)$, $Y_1(x,0;k')$ satisfying (3.6) with $m(x;k)$ and $m(x;k')$, respectively, and $Y_2(l,x;k)$, $Y_2(l,x;k')$ satisfying (3.7) with $m(x;k)$ and $m(x;k')$, respectively. Let Y_{11} be defined by

$$Y_{11}(x,0;k,k') = Y_1(x,0;k) \oplus Y_1(x,0;k'). \tag{9.12}$$

Here $\oplus$ denotes direct sum. From (3.6) it follows that

$$\frac{d}{dx}Y_{11}(x,0;k,k') = \epsilon\hat{m}(x;k,k')Y_{11}(x,0;k,k'), \qquad x \ge 0,$$
$$Y_{11}(0,0;k,k') = Y_g(k) \oplus Y_g(k') \equiv Y_g \oplus Y_g', \tag{9.13}$$
$$\hat{m}(x;k,k') = m(x;k) \oplus m(x;k').$$

Similarly, we define $Y_{22}(l,x;k,k')$ as the direct sum of $Y_2(l,x;k)$ and $Y_2(l,x;k')$ and obtain the equation it satisfies from (3.7). Now we apply the analysis of Sec. 4 to the direct sum processes $Y_{11}(x,0;k,k')$ and $Y_{22}(l,x;k,k')$. First, however, we must introduce some notation.

Let 0 denote the 2×2 zero matrix and define

$$\eta_i^{(1)} = \eta_i \oplus 0, \qquad \eta_i^{(2)} = 0 \oplus \eta_i, \qquad i = 1,2,3, \tag{9.14}$$

where η_i are given by (4.1). Then, $\hat{m}(x;k,k')$ may be expressed as

$$\hat{m}(x;k,k') = \sum_{j=1}^{3} m_j(x;k)\eta_j^{(1)} + \sum_{j=1}^{3} m_j(x,k')\eta_j^{(2)}. \tag{9.15}$$

Here we have employed the notation introduced in (4.3) and we have shown the dependence on k and k' explicitly. Let $Y_{11}^{(\epsilon)}(\sigma;k,k')$ be defined by

$$Y_{11}^{(\epsilon)}(\sigma;k,k') = Y_{11}(\sigma/\epsilon^2,0;k,k'), \qquad \sigma = \epsilon^2 x. \tag{9.16}$$

We are ready now to apply the limit theorem of Sec. 4 to the direct sum process $Y_{11}^{(\epsilon)}(\sigma;k,k')$.

The application of the limit theorem is straightforward because we have arranged that k and k' be distinct. Thus, we find that if f is a bounded smooth function of $Y_1^{(\epsilon)}$ then $U^{(\epsilon)}(\sigma, Y_g \oplus Y_g') = E\{f(Y_{11}^{(\epsilon)}(\sigma))\}$ converges, as $\epsilon \to 0$ and σ remains fixed, to $U^{(0)}(\sigma, Y_g \oplus Y_g')$ where

$$\frac{\partial}{\partial\sigma}U^{(0)} = (V + V' + W)U^{(0)}, \qquad \sigma > 0,$$
$$U^{(0)}(0) = f(Y_g \oplus Y_g'). \tag{9.17}$$

Here V, V', and W are given by

$$V = \alpha(k)(D_{\eta_2^{(1)}}D_{\eta_2^{(1)}} + D_{\eta_3^{(1)}}D_{\eta_4^{(1)}}) + \gamma(k)D_{\eta_1^{(1)}}D_{\eta_1^{(1)}} - \beta(k)D_{\eta_1^{(1)}}, \tag{9.18}$$

$$V' = \alpha(k')(D_{\eta_2^{(2)}}D_{\eta_2^{(2)}} + D_{\eta_3^{(2)}}D_{\eta_3^{(2)}}) + \gamma(k')D_{\eta_1^{(2)}}D_{\eta_1^{(2)}} - \beta(k')D_{\eta_1^{(2)}} \tag{9.19}$$

$$W = 2\delta(k,k')D_{\eta_1^{(1)}}D_{\eta_1^{(2)}}, \tag{9.20}$$

$$\alpha(k) = \frac{k^2}{2}\int_0^\infty R(s;k,k)\cos 2ksds,$$
$$\beta(k) = \frac{k^2}{2}\int_0^\infty R(s;k,k)\sin 2ksds,$$
$$\gamma(k) = k^2\int_0^\infty R(s;k,k)ds,$$
$$\delta(k,k') = \frac{kk'}{2}\int_0^\infty [R(s;k,k') + R(s;k',k)]ds,$$
$$R(s;k,k') = E\{\mu(x;k)\mu(x+s;k')\}. \tag{9.21}$$

Note that W is the interaction operator which couples the statistics of the two components of the direct sum process.

As in Sec. 5 we introduce polar coordinates in order to facilitate application of the limit theorem. Instead of (5.1), however, it is more convenient to use

$$Y = \begin{pmatrix} a & b \\ \bar{b} & \bar{a} \end{pmatrix} = \begin{pmatrix} e^{i\xi}\sqrt{u+1}/\sqrt{2} & e^{i\eta}\sqrt{u-1}/\sqrt{2} \\ e^{-i\eta}\sqrt{u-1}/\sqrt{2} & e^{-i\xi}\sqrt{u+1}/\sqrt{2} \end{pmatrix}, \quad (9.22)$$

$$0 \le \xi < 2\pi, \quad 0 \le \eta < 2\pi, \quad u \ge 1.$$

With this parametrization V of (9.18) and (5.11) takes the form

$$V = \alpha(k)\left[\frac{\partial}{\partial u}\left((u^2-1)\frac{\partial}{\partial u}\right) + \frac{1}{4}\left(\frac{u-1}{u+1}\right)\frac{\partial^2}{\partial \xi^2} + \frac{1}{4}\left(\frac{u+1}{u-1}\right) \times \frac{\partial^2}{\partial \eta^2} + \frac{1}{2}\frac{\partial^2}{\partial \xi \partial \eta}\right] + \frac{\gamma(k)}{4}\left(\frac{\partial}{\partial \xi} + \frac{\partial}{\partial \eta}\right)^2 - \frac{\beta(k)}{2}\left(\frac{\partial}{\partial \xi} + \frac{\partial}{\partial \eta}\right). \quad (9.23)$$

Similarly, V' of (9.19) coincides with (9.23) except α, β, and γ are evaluated at k' and (u, ξ, η) is replaced by (u', ξ', η') which parametrize the second component of the direct sum process. The interaction operator W is given by

$$W = \frac{\delta}{2}\left(\frac{\partial}{\partial \xi} + \frac{\partial}{\partial \eta}\right)\left(\frac{\partial}{\partial \xi'} + \frac{\partial}{\partial \eta'}\right). \quad (9.24)$$

The parametrization (9.22) is more convenient here because, as we will see below, the quantities of interest depend upon all three parameters of Y and they do not simplify as in (5.12), (5.13). Nevertheless other simplifications occur and they are best exploited by using (9.22).

From (3.9) and (3.10) we obtain the expression

$$A(l,x;k)\overline{A}(l,x;k') = E_g \overline{E}'_g \bar{a}_g a'_g \frac{\bar{a}_2 a'_2}{(\bar{b}_2 b_1 + \bar{a}_2 \bar{a}_1)(b'_2 \bar{b}'_1 + a'_2 a'_1)} \quad (9.25)$$

$$B(l,x;k)\overline{B}(l,x;k') = E_g \overline{E}'_g \bar{a}_g a'_g \frac{\bar{b}_2 b'_2}{(\bar{b}_2 b_1 + \bar{a}_2 \bar{a}_1)(b'_2 \bar{b}'_1 + a'_2 a'_1)}. \quad (9.26)$$

Here the primes on the variables indicate that they correspond to k' and the subscripts one and two refer to the direct sum processes $Y_{11}(x, 0; k, k')$ and $Y_{22}(l, x; k; k')$, respectively. The denominators in (9.25) and (9.26) can be expanded into absolutely convergent geometric series. On using, in addition, the parametrization (9.22) we find that

$$A(l,x;k)\overline{A}(l,x;k') = E_g \overline{E}'_g \bar{a}_g a'_g 2 \cdot \sum_{n,m=0}^{\infty} (-1)^{m+n} G^{(1)}_{mn} G^{(2)}_{mn}, \quad (9.27)$$

$$B(l,x;k)\overline{B}(l,x;k') = E_g \overline{E}'_g \bar{a}_g a'_g 2 \cdot \sum_{n,m=0}^{\infty} (-1)^{m+n} G^{(1)}_{mn} G^{(2)}_{(m+1)(n+1)}, \quad (9.28)$$

$$G^{(1)}_{mn}(u_1, \xi_1, \eta_1, u'_1, \xi'_1, \eta'_1) = e^{i[m\eta_1 + (m+1)\xi_1 - n\eta'_1 - (n+1)\xi'_1]} \times \left[\frac{(u_1-1)^{m/2}(u'_1-1)^{n/2}}{(u_1+1)^{(m+1)/2}(u'_1+1)^{(n+1)/2}}\right], \quad (9.29)$$

$$G^{(2)}_{mn}(u_2, \xi_2, \eta_2, u'_2, \xi'_2, \eta'_2) = e^{i[m\xi_2 - m\eta_2 - n\xi'_2 + n\eta'_2]}\left[\frac{(u_2-1)^{m/2}(u'_2-1)^{n/2}}{(u_2+1)^{m/2}(u'_2+1)^{n/2}}\right]. \quad (9.30)$$

Since $G^{(1)}_{mn}$ and $G^{(2)}_{mn}$ are bounded continuous functions of Y_{11} and Y_{22}, respectively, the independent increments property applied to the direct sum processes yields

$$\lim_{\epsilon \to 0} E\{A(\tau/\epsilon^2, \sigma/\epsilon^2; k)\overline{A}(\tau/\epsilon^2, \sigma/\epsilon^2; k')\} = E_g \overline{E}'_g \bar{a}_g a'_g 2 \cdot \sum_{n,m=0}^{\infty} (-1)^{m+n} \hat{G}^{(1)}_{mn}(\sigma)\hat{G}^{(2)}_{mn}(\tau - \sigma), \quad (9.31)$$

$$\lim_{\epsilon \to 0} E\{B(\tau/\epsilon^2, \sigma/\epsilon^2; k)\overline{B}(\tau/\epsilon^2, \sigma/\epsilon^2; k')\} = E_g \overline{E}'_g \bar{a}_g a'_g 2 \cdot \sum_{n,m=0}^{\infty} (-1)^{m+n} \hat{G}^{(1)}_{mn}(\sigma)\hat{G}^{(2)}_{(m+1)(n+1)}(\tau - \sigma) \quad (9.32)$$

Here $\hat{G}^{(1)}_{mn}(\sigma)$ and $\hat{G}^{(2)}_{mn}(\sigma)$ are functions of the generator and load parameters, respectively, and they satisfy the initial value problems

$$\frac{\partial}{\partial \sigma}\hat{G}^{(1)}_{mn} = (V + V' + W)\hat{G}^{(1)}_{mn}, \quad \sigma > 0, \quad \hat{G}^{(1)}_{mn}(0) = G^{(1)}_{mn} \quad (9.33)$$

$$\frac{\partial}{\partial \sigma}\hat{G}^{(2)}_{mn} = (V + V' + W)\hat{G}^{(2)}_{mn}, \quad \sigma > 0, \quad \hat{G}^{(2)}(0) = G^{(2)}_{mn}. \quad (9.34)$$

The solution of (9.33) and (9.34) is obtained easily after observing that the functions

$$Q_{MN}(u, \xi, \eta) = (u-1)^{M/2}(u+1)^{-N/2}e^{i[\kappa_1 M\eta + \kappa_2 N\xi]}, \quad M, N \ge 0, \quad \kappa_1, \kappa_2 = \pm 1. \quad (9.35)$$

and formal eigenfunctions of V in (9.23). The corresponding eigenvalues are

$$\Lambda_{MN} = \frac{\alpha}{2}(M - N - MN(1 + \kappa_1\kappa_2)) - \frac{\gamma}{4}(M + \kappa_1\kappa_2 N)^2 + \frac{i\beta}{2}(\kappa_1 M + \kappa_2 N). \quad (9.36)$$

From (9.29), (9.30), and (9.35) it follows that both $G^{(1)}_{mn}$ and $G^{(2)}_{mn}$ are formal eigenfunctions of $V + V' + W$ in (9.17) with eigenvalues $\lambda^{(1)}_{mn}$ and $\lambda^{(2)}_{mn}$ given by

$$\lambda^{(1)}_{mn} = -\alpha(k)(m^2 + m + \tfrac{1}{2}) - \gamma(k)(m + \tfrac{1}{2})^2 - \alpha(k')(n^2 + n + \tfrac{1}{2}) - \gamma(k')(n + \tfrac{1}{2})^2 + 2\delta(m + \tfrac{1}{2})(n + \tfrac{1}{2}) + i[-\beta(k)(m + \tfrac{1}{2}) + \beta(k')(n + \tfrac{1}{2})], \quad (9.37)$$

$$\lambda^{(2)}_{mn} = 0. \quad (9.38)$$

Thus

$$\hat{G}^{(1)}_{mn} = e^{\lambda^{(1)}_{mn}\sigma}G^{(1)}_{mn}, \quad \hat{G}^{(2)}_{mn} = G^{(2)}_{mn}. \quad (9.39)$$

The fact that $G^{(2)}_{mn}$ is independent of σ, a consequence of (9.38), is remarkable. It implies that the limits (9.31) and (9.32) are independent of τ and, after some rearrangement, are given by

$$\lim_{\epsilon \to 0} E\{A(\tau/\epsilon^2, \sigma/\epsilon^2; k)\overline{A}(\tau/\epsilon^2, \sigma/\epsilon^2; k')\} = E_g \overline{E}'_g \cdot \sum_{n,m=0}^{\infty} e^{\lambda^{(1)}_{mn}\sigma}[\Gamma_g(k)\Gamma_l(k)]^m[\overline{\Gamma}_g(k')\overline{\Gamma}_l(k')]^n, \quad (9.40)$$

$$\lim_{\epsilon \to 0} E\{B(\tau/\epsilon^2, \sigma/\epsilon^2; k)\overline{B}(\tau/\epsilon^2, \sigma/\epsilon^2; k')\} = \Gamma_l(k)\overline{\Gamma}_l(k')E_g \overline{E}'_g \cdot \sum_{n,m=0}^{\infty} e^{\lambda^{(1)}_{mn}\sigma}[\Gamma_g(k)\Gamma_l(k)]^m \times [\overline{\Gamma}_g(k')\overline{\Gamma}_l(k')]^n. \quad (9.41)$$

Here $\Gamma_g(k)$ and $\Gamma_l(k)$ are as defined in (5.3).

The results (9.40) and (9.41) indicate that the interesting phenomena of power transfer due to fluctuation,

described in Secs. 6 and 7, do not occur for pulses of finite total power (cf. (9.8)). The total power must be at least of order $1/\epsilon^2$ before cumulative phenomena due to fluctuation become significant. For example, when the load is matched so that $\Gamma_l = 0$, (9.41) and (9.7) show that within our approximation, there is no power reflected. This should be contrasted with the results of Secs. 6 and 7. The analysis of the pulse problem when (9.8) is not valid is considerably more complicated and will not be considered here.

In the remainder of this section we apply the above analysis to the following problem:

$$\mathcal{E}_g(t) = e^{-t^2/2s^2}\cos\omega_0 t, \tag{9.42}$$

$$\Gamma_g = \Gamma_l = 0, \tag{9.43}$$

$$\alpha(k) = \frac{k^2}{2}, \quad \beta(k) = \frac{k^2}{2}, \quad \gamma(k) = k^2, \quad \delta(k,k') = kk'. \tag{9.44}$$

The generator voltage is a Gaussian modulated pulse (9.42), the line is matched to both the generator and the load and the spectrum of the random process $\mu(x)$ is assumed flat (9.44). Both (9.42) and (9.44) violate our previous hypotheses which are convenient idealizations.

From (9.1) and (9.42) we obtain $e_g(\omega)$:

$$e_g(\omega) = \frac{s}{2}(e^{-(s^2/2)(\omega-\omega_0)^2} + e^{-(s^2/2)(\omega+\omega_0)^2} \tag{9.45}$$

From (9.43), (9.44) and (9.40), (9.37) we have

$$\lim_{\epsilon\to 0} E\{A(\tau/\epsilon^2, \sigma/\epsilon^2; k)\overline{A}(\tau/\epsilon^2, \sigma/\epsilon^2; k')\} = [e_g(\omega)\bar{e}_g(\omega')/4Z_0]e^{-(1/2)(k^2+k'^2-kk')\sigma}. \tag{9.46}$$

Note again that (9.46) is independent of τ. On inserting (9.46) in (9.6) and performing the double integral we find that when $\epsilon \ll 1$ and $x \gg 1$ so that $\epsilon^2 x = \sigma$ then,

$$E\{|\mathcal{A}(l,x;t)|^2\} \sim \frac{s^2}{8Z_0}\left[\left(s^2 + \frac{3\epsilon^2 x}{2c^2}\right)\left(s^2 + \frac{\epsilon^2 x}{2c^2}\right)\right] \times \exp\left(\frac{-(t-x/c)^2}{2s^2 + 3\epsilon^2 x/c^2}\right)\left[\exp\left(\frac{-\omega_0^2 s^2 \epsilon^2 x}{2c^2(s^2 + \epsilon^2 x/2c^2)}\right) + \exp\left(\frac{-3\omega_0^2 s^2 \epsilon^2 x}{2c^2(s^2 + 3\epsilon^2 x/2c^2)}\right)\cos\left(\frac{2\omega_0 s(t-x/c)}{s^2 + 3\epsilon^2 x/2c^2}\right)\right]. \tag{9.47}$$

By comparing $\mathcal{E}_g^2(t)$ in (9.42) and (9.47) we obtain the pulse spreading factor: $1 + 3\epsilon^2 x/2c^2 s^2$.

ACKNOWLEDGMENT

We wish to thank J. B. Keller for several discussions on the subject of this paper and for reading the manuscript and suggesting improvements.

*The research for this paper was supported in part by the Office of Naval Research under Contract No. N00014-67-A-0467-0015 and by the Air Force Office of Scientific Research under Grant No. AFOSR-71-2013.

† Present address: Dept. of Math., Virginia Polytechnic Inst. & State U., Blacksburg, Va.

[1]R. Z. Khasminskii, Theor. Probability Appl. **11,** 390 (1966).
[2]G. C. Papanicolaou and S. R. S. Varadhan, Comm. Pure Appl. Math. (to appear).
[3]Yu. L. Gazaryan, Zh. Eksp. Teor. Fiz. **56,** 1856 (1969) [Sov. Phys.-JETP **29,** 996 (1969)].
[4]R. H. Lang, J. Math. Phys. (to appear).
[5]A. Schuster, Astrophys. J. **21,** 1 (1905).
[6]S. Chandrasekhar, *Radiative Transfer* (Oxford, U. P., London, 1950).
[7]D. Marcuse, IEEE Trans. Microwave Theory Tech. **20,** 541 (1972).
[8]M. S. Howe, Phil. Trans. Roy. Soc. London **274,** 523 (1973).
[9]I. M. Besieris and F. D. Tappert, J. Math. Phys. **13,** 704 (1973).
[10]F. P. Kapron, R. D. Mauer, and M. P. Teter, Appl. Opt. **11,** 1352 (1972).
[11]G. C. Papanicolaou, J. Math. Phys. **13,** 1912 (1972).
[12]K. M. Watson, Phys. Fluids **13,** 2514 (1970); J. Math. Phys. **10,** 688 (1968).
[13]C. W. Lau and K. M. Watson, J. Math. Phys. **11,** 3125 (1970).
[14]J. L. Peacher and K. M. Watson, J. Math. Phys. **11,** 1496 (1970).
[15]J. A. Morrison, J. Math. Anal. Appl. **39,** 13 (1972).
[16]U. Frisch, in *Probabilities Methods in Applied Mathematics*, edited by A. T. Bharucha-Reid (Academic, New York, 1968).
[17]J. A. Morrison and J. McKenna, *Proceedings of A. M. S. Symposium on Stochastic Equations, March 1972* (to appear).
[18]S. Helgason, in *Battelle Rencontres*, edited by C. De Witt (Benjamin, New York, 1968).
[19]R. Burridge and G. C. Papanicolaou, Commun. Pure Appl. Math. **25,** 715 (1972).
[20]W. Magnus, F. Oberhettinger, and R. P. Soni, *Formulas and Theorems for the Special Functions of Mathematical Physics* (Springer, New York, 1966).
[21]P. L. Sulem and U. Frisch, J. Plasma Phys. **8,** 217 (1972).
[22]P. L. Sulem, preprint, Observatoire de Nice.
[23]U. Frisch, C. Froeschle, J.-P. Scheidecker, and P. L. Sulem, preprint, Observatoire de Nice.
[24]J. Morrison, G. C. Papanicolaou, and J. B. Keller, Commun. Pure Appl. Math. **24,** 473 (1971).
[25]G. C. Papanicolaou, SIAM J. Appl. Math. (Soc. Ind. Appl. Math.) **21,** 13 (1971).
[26]G. C. Papanicolaou and J. B. Keller, SIAM J. Appl. Math. (Soc. Ind. Appl. Math.) **21,** 287 (1971).
[27]V. V. Sobolev, *A Treatise on Radiative Transfer* (Van Nostrand, Princeton, 1963).
[28]J. A. Morrison, IEEE Trans. MTT (to appear).

Time evolution of a two-dimensional model system. I. Invariant states and time correlation functions

J. Hardy and Y. Pomeau

Laboratoire de Physique des Plasmas, Faculté des Sciences, 91-Orsay, France

O. de Pazzis

Centre de Physique Théorique, C.N.R.S., Marseille, France
(Received 3 January 1973)

This paper is the first one of a series devoted to the study of a particularly simple two-dimensional system of classical particles. The model is presented and some general feautres of it are established. We prove that, among states without correlations between particles with different velocities, there is a unique time invariant state with given density and hydrodynamic velocity. This "equilibrium state" is studied in detail. In particular its ergodic and mixing properties are investigated. We propose an approximation in order to estimate the asymptotic part of the time correlation functions and show that the long time tail is ruled by the "hydrodynamic" behavior of the model, namely by the evolution of the long wavelength perturbations.

1. INTRODUCTION

The aim of this paper is to propose a somewhat simple system of classical particles with interactions and to investigate some of its properties. As an introduction we would like to give some explanation of the motivation which leads us to undertake such a work.

The process of approach to equilibrium of macroscopic systems is, since Boltzmann's work, an old problem. It is known that this question is closely related to the ergodic hypothesis. In fact it can be proved that if the ergodic hypothesis holds, then every local perturbation of the equilibrium relaxes in a well-defined sense. However, the ergodic hypothesis is not sufficient to lead to a full understanding of transport phenomena in fluids and gases. In fact the ergodic hypothesis predicts the relaxation of time correlation function, but it does not give any information about the exact asymptotic behavior of the correlation functions. Still the knowledge of the long time tail of suitable correlation functions is of great importance in order to ensure the existence of transport coefficients as given by Kubo's formulas.

As far as infinite systems are concerned there is no realistic model for which the ergodic hypothesis and existence of transport coefficients have both been investigated. In one dimension, the model of hard rods has been shown to have very good properties of mixing.[1,2] Furthermore, it is possible to calculate exactly the self-diffusion coefficient for this model.[3] Unfortunately, the behavior of the hard rods model is far from being that of realistic systems; in particular the process of thermalization is absent because any velocity distribution function is an equilibrium one. In a general way, one-dimensional systems have undoubtedly very particular properties which cannot be expected to occur in two- or three-dimensional systems.

When considering infinite systems, in two or three dimensions, one is immediately faced with the complicated problem of time evolution of configurations with infinitely many particles. In fact it is easy to construct configurations of hard discs or hard spheres, for instance, leading in a finite time to a catastrophic situation. The problem is to prove that the probability of occurrence of these "bad" configurations is zero at equilibrium.[4] Although this problem has been solved for hard squares and hard cubes with discrete velocities,[5] this last model is still too complicated for investigating in a rigorous way the ergodic hypothesis and the existence of transport coefficients. The same remark holds for the so-called Maxwell model, for which a hydrodynamic theory has recently been developed,[6] and the divergence of the viscosity coefficient has been discussed in detail.[7]

Here we propose a two-dimensional model of particles with discrete velocities and discrete positions; namely, the particles are on the sites of a two-dimensional lattice and jump each unit of time in one of the four directions of the lattice suffering collisions with each other in such a way that the particle number and the total momentum are conserved during a collision. There is no difficulty in the definition of time evolution of infinite configurations and, as it will be seen, some interesting results can be obtained with this model. The authors do not pretend to have imagined a model which will play the same important role as the Ising model does in the theory of phase transitions, but the results related in this paper seemed to them sufficiently stimulating and they hope to arrive in the future at a more profound understanding of the relaxation processes with this model.

In Sec. 2 we describe the model in detail and define the time evolution mapping T, on the space of infinite configurations. Some elementary and simple properties of T are quoted in Sec. 3. The nonreversible character of T is emphasized and a slight modification is proposed in order to preserve the microreversibility. Section 4 is devoted to the search for invariant states. In particular a uniqueness theorem is proved on a restricted class of homogeneous states. In Sec. 5 we recall the definition of "mixing" systems and show that for the invariant states studied here the mixing property is equivalent to the asymptotic relaxation of a reduced class of time correlation functions. In Sec. 6 we propose an approximation of the evolution equations which permits us to compute the most simple correlation functions and to give their asymptotic behavior as the time tends to infinity. Finally we discuss the results obtained in Sec. 6 and the meaning of the approximation we have used; in particular we show that this approximation can be used to prove in a rigorous way the existence of sound waves propagation in this system.

2. DESCRIPTION OF THE MODEL

We consider an infinite square lattice. On each lattice site there are at most four particles. The velocity of a particle is one of the four unit vectors (1, 0), (0, 1),

$(-1,0)$, or $(0,-1)$, which we label respectively with the four numbers $(1,2,3,4)$. The configurations where there are at least two particles with the same velocity on the same lattice site are excluded. We denote by X an allowed configuration and by K the set of all X. It will be convenient to introduce the following functions defined in K:

$$\sigma_X(p,q;i) = \begin{cases} 1 & \text{if the site } (p,q) \text{ is occupied by a particle of } X \text{ of velocity } i, \\ 0 & \text{otherwise.} \end{cases}$$

Giving an X is the same as giving the values of all the functions

$$\sigma_X(p,q;i), \text{ where } (p,q) \in \mathbb{Z}^2 \text{ and } i \in P = \{1,2,3,4\}.$$

We may write then $K = \{0,1\}^{4\mathbb{Z}^2}$ and from Tychonov's theorem K is a compact set. We shall need in the subsequent sections the following natural decomposition of K; namely, if we denote by X_i where $i \in P$ the subconfiguration of X formed with the particles of velocity i, it results from the exclusion principle that X_i is an element of $K_i = (0,1)^{\mathbb{Z}^2}$. Therefore, K has the structure of a product: $K = \prod_{i=1}^{4} K_i$. A "state" on K will be a probability measure on K, that is, a positive measure ρ such that $\rho(K) = 1$.

Let us now introduce the time evolution of a configuration. The movement of the particles is specified as follows. During one unit of time each particle jumps one step in the direction of its velocity; then on each lattice site where one of the two situations indicated in Fig. 1 occurs, there is a collision exchanging the situations (a) and (b) with each other. On the other lattice sites the situation is left unchanged. In Fig. 2 we give an example of a collision at time t, and the corresponding configurations at times $t-1, t+1$. Note that it is not merely a zero-impact parameter model since a collision in the middle of the line connecting adjacent sites is not allowed. If X is the initial configuration, we call $T(X)$ the new configuration after one unit of time has elapsed. It is clear that $T(X)$ is still an allowed configuration and that T maps K onto itself homeomorphically. This last property which does not hold for realistic systems[4,5] eliminates some inessential and purely technical difficulties.

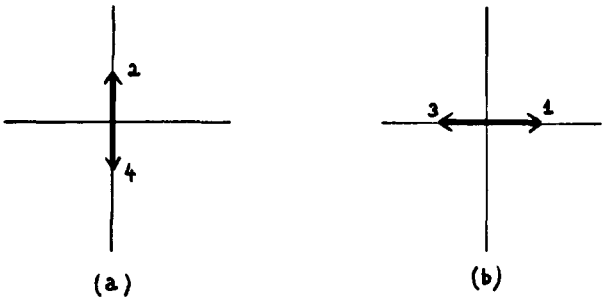


FIG. 1. The two situations giving a collision.

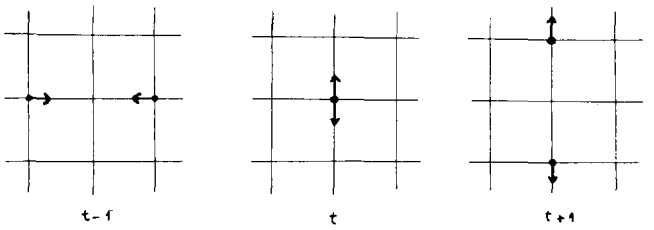


FIG. 2. An example of a collision at time t, and the corresponding configurations at times $t-1, t+1$.

The time evolution mapping T can be written with the functions $\sigma_X(p,q;i)$ in the following way:

$$\begin{aligned} \sigma_{T(X)}(p,q;1) &= \sigma_X(p-1,q;1) - \psi_X(p,q), \\ \sigma_{T(X)}(p,q;2) &= \sigma_X(p,q-1;2) + \psi_X(p,q), \\ \sigma_{T(X)}(p,q;3) &= \sigma_X(p+1,q;3) - \psi_X(p,q), \\ \sigma_{T(X)}(p,q;4) &= \sigma_X(p,q+1;4) + \psi_X(p,q), \end{aligned} \tag{2.1}$$

where

$$\begin{aligned} \psi_X(p,q) = {} & \sigma_X(p-1,q;1)\sigma_X(p+1,q;3)\bar{\sigma}_X(p,q-1;2) \\ & \times \bar{\sigma}_X(p,q+1;4) - \bar{\sigma}_X(p-1,q;2)\bar{\sigma}_X(p+1,q;3) \\ & \times \sigma_X(p,q-1;2)\sigma_X(p,q+1;4) \end{aligned}$$

and

$$\bar{\sigma}_X(p,q;i) = 1 - \sigma_X(p,q;i).$$

Before giving some elementary properties of T, we make two remarks: First the particle number and the total momentum are conserved during a collision, and second the impact parameter is always zero in a collision so that on each horizontal (resp. vertical) line of lattice sites the horizontal (resp. vertical) momentum is conserved. This peculiarity will lead to specific results when we consider relaxation of time correlation functions.

3. ELEMENTARY PROPERTIES OF T

First note that T splits into two mappings: $T = C \cdot T_0$, where T_0 and C are respectively the free evolution mapping and the collision operator. The operator C can be written in the following way:

$$\sigma_{C(X)}(p,q;i) = \sigma_X(p,q;i) + (-)^i \psi_{T_0^{-1}(X)}(p,q) \tag{3.1}$$

The mapping C is strictly local and is a symmetry mapping: $C^2 = 1$. Obviously C and T_0 do not commute. The above decomposition is a peculiarity of the model, in realistic systems the unitary evolution operator reads as $U(t) = \exp(-itH_0 - itH^1)$ and cannot split into $\exp(-itH_0) \times \exp(-itH^1)$ because $[H_0, H^1] \neq 0$. It is only *the infinitesimal generator* which splits into a free part on a collision part.

We would like to emphasize the fact that although T is one to one, it does not satisfy the microreversibility principle. More precisely let us introduce the mapping $\mathcal{R}: K \to K$ which consists in reversing the velocities of the particles:

$$\sigma_{\mathcal{R}(X)}(p,q;i) = \sigma_X(p,q;i+2)$$

An elementary calculation shows that $\mathcal{R} \cdot T^{-1} \neq T \cdot \mathcal{R}$ and the reversibility principle is violated. Still there is a slight modification of the evolution mapping T which does lead to the reversibility principle.[8] The trick is to take as the evolution mapping $\bar{T} = T_0 \cdot C \cdot T_0$ which amounts to look at the system at even times only and to have collisions at odd times only. It is easy to show that $\mathcal{R} \cdot \bar{T}^{-1} = \bar{T} \cdot \mathcal{R}$. In addition one can convince oneself that all the results of this paper are valid for $\bar{T}$ after some trivial modifications. Therefore we prefer to investigate the first mapping T because the algebra in this case is a little bit simpler.

Another symmetry which will play an interesting role is the symmetry between occupied and empty sites.

Namely let us introduce the mapping $\mathcal{L}$: $K \to K$ such that $\sigma_{\mathcal{L}(X)}(p,q;i) = \bar{\sigma}_X(p,q;i)$. It is easy to verify that T commutes with $\mathcal{L}$. In other words any dynamical property valid in the range of low density has a corresponding property in the range of high density.

We conclude this section with some considerations about the action of translations. It is clear that $\mathbb{Z}^2$ acts in a natural way on K, if g belongs to $\mathbb{Z}^2$ let us denote by τ_g its action on K, one has immediately:

$$\tau_g \cdot T = T \cdot \tau_g \tag{3.2}$$

A state ρ which is invariant under translations will be called homogeneous. If ρ is homogeneous we can define its entropy.[9]

$$H(\rho) = \lim_{\Lambda\to\infty} \frac{1}{4N(\Lambda)} \left(\sum_{X\subset\Lambda} - \rho_\Lambda(X) \log \rho_\Lambda(X) \right), \tag{3.3}$$

where Λ is a finite subset of $\mathbb{Z}^2$ with $N(\Lambda)$ sites and $\rho_\Lambda(X)$ means the probability of having the configuration X in Λ. The extra factor 4 is due to the possibility of having four particles, on the same lattice site. It is useful to note that $H(\rho)$ is the entropy of the dynamical system $\{K, \mathbb{Z}^2, \rho\}$ as defined in Ref. 10. It results from this remark and from (3.2) that $H(\rho \cdot T) = H(\rho)$ which means that the entropy is conserved during time evolution. For completeness we give a direct proof of this last property using the definition (3.3).

Proposition 3.1: If ρ is a homogeneous measure on K then $H(\rho \cdot T) = H(\rho)$.

Proof: For convenience we choose in the definition (3.2) the particular sequence of subsets Λ_n, where Λ_n is the square centered at the origin with its diagonals of length $2n$ lying on the $0x$ and $0y$ axis. In this case

$$(\rho \cdot T)_{\Lambda_n}(X) = \sum_{\substack{y\subset\Lambda_{n+1} \\ T(y)\cap\Lambda_n = X}} \rho_{\Lambda_{n+1}}(y);$$

therefore,

$$-(\rho \cdot T)_{\Lambda_n}(X) \log (\rho \cdot T)_{\Lambda_n}(X) \leq - \sum_{\substack{y\subset\Lambda_{n+1} \\ T(y)\cap\Lambda_n = X}} \rho_{\Lambda_{n+1}}(y) \log \rho_{\Lambda_{n+1}}(y)$$

and

$$H(\rho \cdot T) \leq \lim_{n\to\infty} - \frac{1}{4N(\Lambda_n)} \sum_{X\subset\Lambda_n} \sum_{\substack{y\subset\Lambda_{n+1} \\ T(y)\cap\Lambda_n = X}} \rho_{\Lambda_{n+1}}(y) \log \rho_{\Lambda_{n+1}}(y)$$

$$\leq \lim_{n\to\infty} - \frac{1}{4N(\Lambda_n)} \sum_{y\subset\Lambda_{n+1}} \rho_{\Lambda_{n+1}}(y) \log \rho_{\Lambda_{n+1}}(y) = H(\rho).$$

In the same way $H(\rho \cdot T^{-1}) \leq H(\rho)$; therefore, $H(\rho \cdot T) = H(\rho)$.

Note that although T is not strictly reversible, it satisfies a principle of conservation of entropy so that with regard to statistical mechanics there is no special difference between the past and the future.

4. INVARIANT STATES

This section is devoted to the study of a particular class of states invariant under T. In the remainder of this paper we shall consider homogeneous states only. If ρ is a homogeneous state, we can define the four quantities $n_i(\rho) = \int \sigma_X(p,q;i)d\rho$, that are independent on (p,q). From these four quantities we can construct the density $n(\rho) = \sum_{i=1}^4 n_i(\rho)$ and the hydrodynamic momentum $n(\rho)v_x(\rho) = n_1(\rho) - n_3(\rho)$, $n(\rho)v_y(\rho) = n_2(\rho) - n_4(\rho)$. It is natural to ask the question: Given (n, v_x, v_y), does there exist a homogeneous state ρ such that $n(\rho) = n$, $v_x(\rho) = v_x$, and $v_y(\rho) = v_y$?

Proposition 4.1: Given $(n, v_x, v_y) \in \mathbb{R}^3$, there exists at least one homogeneous state ρ such that $n(\rho) = n$, $v_x(\rho) = v_x$, $v_y(\rho) = v_y$ iff

$$\begin{aligned} &0 \leq n \leq 4, \\ &|nv_x| \leq 1, \quad |nv_y| \leq 1, \\ &|nv_x| + |nv_y| \leq \min(n; 4-n). \end{aligned} \tag{4.1}$$

Proof: First let us show that condition (4.1) is a necessary one. In fact let ρ be a solution of the problem; thus $0 \leq n_i(\rho) \leq 1$ $\forall i \in P$. We have then the system of inequalities:

$$\begin{aligned} &0 \leq n_1 = n/4 + nv_x/2 + \chi \leq 1, \\ &0 \leq n_2 = n/4 + nv_y/2 - \chi \leq 1, \\ &0 \leq n_3 = n/4 - nv_x/2 + \chi \leq 1, \\ &0 \leq n_4 = n/4 - nv_y/2 - \chi \leq 1, \end{aligned} \tag{4.2}$$

where $4\chi = \sum_{i=1}^4 (-)^{i+1} n_i$. Now it is clear that the existence of a χ satisfying (4.2) is equivalent to

$$\max(-n/4 + |nv_x|/2, n/4 + |nv_y|/2 - 1) \leq \min(n/4 - |nv_y|/2, -n/4 - |nv_x|/2 + 1).$$

The above inequality is equivalent to (4.1). Conversely, if (4.1) holds, there exists at least one χ satisfying (4.2) and there exist n_i, $i \in P$ such that $0 \leq n_i \leq 1$ and $\sum_{i=1}^4 n_i = n$, $nv_x = n_1 - n_3$, $nv_y = n_2 - n_4$. It suffices then to take the state defined by $\int \Pi_{i,p,q} \sigma_X(p,q;i)d\rho = \Pi_{i,p,q} n_i$ for any finite product of different functions $\sigma_X(p,q;i)$, to have a solution of the problem.

Let us denote by Δ the open convex bounded set of $\mathbb{R}^3$ where the inequalities (4.1) are strictly satisfied. We have drawn in Fig. 3 some representative sections of Δ by planes $n = cte$. It will be useful to note that boundary of Δ corresponds to the case where one of the four

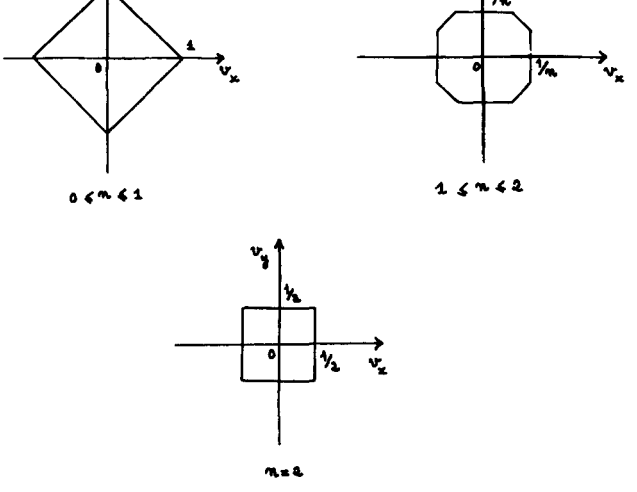


FIG. 3. Typical sections of the domain Δ by planes $n = cte$.

numbers n_i is equal to 0 or 1. (The converse is not true.)

In view of the time evolution of the model the introduction of the hydrodynamic quantities, (n, v_x, v_y) is quite natural. In fact we have:

Proposition 4.2: The hydrodynamic quantities $n(\rho)$, $v_x(\rho), v_y(\rho)$ are invariant under T.

Proof: This is an immediate consequence of the fact that the collisions preserve the particle number and the total momentum. Let us prove for example that $n(\rho \cdot T) = n(\rho)$. We have

$$\begin{aligned} n(\rho \cdot T) &= \int \sum_{i=1}^{4} \sigma_X(0,0;i)d\rho \cdot T = \int \sum_{i=1}^{4} \sigma_{T(X)}(0,0;i)d\rho \\ &= \int [\sigma_X(-1,0;1) + \sigma_X(0,-1;2) \\ &\quad + \sigma_X(1,0;3) + \sigma_X(0,1;4)]d\rho \qquad \text{from (2.1)} \\ &= n(\rho) \qquad \text{from the homogeneity condition.} \end{aligned}$$

As the hydrodynamic parameters (n, v_x, v_y) are T-invariant quantities, it is natural to ask whether there is a unique T-invariant state such that $n(\rho) = n$, $v_x(\rho) = v_x$, $v_y(\rho) = v_y$. It is expected that, to have a uniqueness property, one must add some conditions on the range of the correlation functions of the states considered. Let us make precise the condition that we have to impose

We denote by $\{K, \mathfrak{A}, \rho\}$ the measure space K, where $\mathfrak{A}$ is the algebra of measurable subsets of K with respect to ρ. We have seen that $\mathbb{Z}^2$ acts on K. If $g \in \mathbb{Z}^2$, let τg be the translation corresponding to g. The weakest condition that we can impose on the range of the correlation functions of the state ρ is

$$\lim_{\Lambda \to \infty} \frac{1}{N(\Lambda)} \sum_{g \in \Lambda} \rho(A \cap \tau_g B) = \rho(A)\rho(B), \tag{4.3}$$

where Λ is for instance a square centered at the origin and $A, B \in \mathfrak{A}$. A state satisfying (4.3) for any pair $A, B \in \mathfrak{A}$ is said to be $\mathbb{Z}^2$-ergodic. We shall show in this section that condition (4.3) is not sufficient to ensure the uniqueness of a T-invariant state with given hydrodynamic parameters. We shall need a stronger condition which is the following:

$$\lim_{|g| \to \infty} \rho(A \cap \tau_g B) = \rho(A)\rho(B) \quad \forall A, B \in \mathfrak{A}. \tag{4.4}$$

It seems to be reasonable to conjecture that condition (4.4) ensures the uniqueness of a T-invariant state with given hydrodynamic parameters. Actually we are not able to prove this strong conjecture. We shall restrict our considerations about states satisfying an additional condition on their correlation functions:

Definition: Let $(n, v_x, v_y) \in \Delta$, we denote by $\mathfrak{F}(n, v_x, v_y)$ the set of homogeneous states ρ satisfying (4.4) with hydrodynamic parameters (n, v_x, v_y) and which are compatible with the decomposition $K = \Pi_{i=1}^4 K_i$, that is $\rho = \Pi_{i=1}^4 \otimes \rho_i$, where ρ_i is a probability measure on K_i.

We emphasize the fact that the factorization condition is introduced for simplification only and is based by no means on physical arguments. We can then prove:

Theorem 4.1: There exists in $\mathfrak{F}(n, v_x, v_y)$ a unique T-invariant state. This state is fully described by the following properties: Each factor ρ_i is a measure without any correlations between lattice sites and the four numbers $n_i = \int \sigma_X(p, q; i)d\rho$ satisfy the equation

$$n_1 n_3 (1 - n_2)(1 - n_4) = n_2 n_4 (1 - n_1)(1 - n_3). \tag{4.5}$$

Proof: Let us give a list of the different steps of the proof.

(a) If $\rho = \Pi_{i=1}^4 \otimes \rho_i$ is homogeneous, then T-invariance is equivalent to C-invariance.

(b) If $\rho = \Pi_{i=1}^4 \otimes \rho_i$ is C-invariant, then (4.5) holds.

(c) There exists a unique system of numbers n_i, $i \in P$, such that

$$\begin{aligned} &0 \leq n_i \leq 1, \\ &n_1 n_3 (1 - n_2)(1 - n_4) = n_2 n_4 (1 - n_1)(1 - n_3), \\ &n = \sum_{i=1}^{4} n_i, \quad v_x = n_1 - n_3, \quad v_y = n_2 - n_4. \end{aligned} \tag{4.6}$$

(d) If $\rho \in \mathfrak{F}(n, v_x, v_y)$ is C-invariant, then each ρ_i is without any correlation between lattice sites.

(e) The properties (c) and (d) characterize a unique state $\rho(n, v_x, v_y)$ which is shown to be T-invariant.

To prove (a), note that if $\rho = \Pi_{i=1}^4 \otimes \rho_i$ is homogeneous, then each ρ_i is homogeneous also, so that $\rho \cdot T_0 = \rho$. Therefore $\rho \cdot T = \rho$ is equivalent to $\rho \cdot C = \rho$. This last equation is simpler because C operates on each lattice site (this is the chief reason why we introduced the factorisation condition on ρ).

To prove (b), we use the function $\Phi_X(p, q) = \Psi_{T_0^{-1}(X)}(p, q)$ introduced in Sec. 2. It is clear that $\Phi_{C(X)}(p, q) = -\Phi_X(p, q)$ so that C-invariance of ρ implies

$$\int \Phi_X(p, q)d\rho = 0.$$

But this equation is exactly (4.5) if $\rho = \Pi_{i=1}^4 \otimes \rho_i$. In order to prove (c), we introduce the parameter χ as in the proof of Proposition 4.1. In terms of the quantities (n, v_x, v_y, χ) Eq. (4.5) reads as follows:

$$\begin{aligned} &(n/4 + nv_x/2 + \chi)(n/4 - nv_x/2 + \chi)(1 - n/4 - nv_y/2 + \chi) \\ &\quad \times (1 - n/4 + nv_y/2 + \chi) \\ &= (n/4 + nv_y/2 - \chi)(n/4 - nv_y/2 - \chi) \\ &\quad \times (1 - n/4 - nv_x/2 - \chi)(1 - n/4 + nv_y/2 - \chi). \end{aligned} \tag{4.7}$$

Now proving (c) consists in proving that if $(n, v_x, v_y) \in \Delta$, there exists a unique χ satisfying (4.7) such that inequalities (4.2) hold. But in the domain of validity of (4.2) the left-hand side of (4.7) is an increasing function of χ whereas the right-hand side is a decreasing one. Therefore, the solution, if it exists, is unique. The existence of a solution is based on a similar argument. In fact, if $(n, v_x, v_y) \in \Delta$, the greatest root of the left-hand side of (4.7) is lower than the lowest root of the right-hand side of (4.7) so that a solution exists and is unique.

The proof of (d) is more intricate and will be decomposed into three lemmas. In the following $a_i = (p_i, q_i)$ will represent a lattice site. Let us introduce the correlation functions

$$\rho_i^{(r)}(a_1, a_2, \ldots, a_r) = \int \sigma_X(a_1; i) \ldots \sigma_X(a_r; i)d\rho,$$

where $\{a_i\}$, $i \in \{1, 2, \ldots, r\}$, is a family of distinct lattice sites. We have to show that $\rho_i^{(r)} = (n_i)^r \quad \forall i \in P$, $\forall r \in \mathbb{N}$. The proof goes by induction. First of all we prove the property for $r = 2$, then for $r = 2l + 1$ with $l \geq 1$, and finally for $r = 2l$ with $l \geq 2$.

Lemma 4.1: If $\rho = \prod_{i=1}^{4} \otimes \rho_i$ is C-invariant and satisfies (4.4), then $\rho_i^{(2)}(a_1, a_2) = (n_i)^2$.

Proof: We start with Eq. (3.1). The C-invariance of ρ implies that

$$\int \sigma(a;1)\Phi(b)d\rho = -\int \sigma(a;2)\Phi(b) = \int \sigma(a;3)\Phi(b)d\rho = -\int \sigma(a;4)\Phi(b)d\rho. \tag{4.8}$$

In these equations we have omitted the subscript X for simplicity. Now if $\rho = \prod_{i=1}^{4} \otimes \rho_i$ and if (4.6) holds, (4.8) reads as

$$A_1 g_1^{(2)}(a,b) = A_2 g_2^{(2)}(a,b) = A_3 g_3^{(2)}(a,b) = A_4 g_4^{(2)}(a,b), \tag{4.9}$$

where $A_i = n_{i+1}n_{i+3}(1 - n_{i+2})/n_i$ and $g_i^{(2)}(a,b) = \rho_i^{(2)}(a,b) - n_i^2$. It is shown in Appendix A that if $(n, v_x, v_y) \in \Delta$, we have $0 < A_i < 1 \ \forall\, i \in P$.

From (3.1) we can also construct the equation

$$\sigma_{C(X)}(a;1)\sigma_{C(X)}(a;3) = \sigma_X(a;1)\sigma_X(a;3) - \Phi_X(a)$$

so that the C-invariance of ρ implies also that

$$\int \sigma(a;1)\Phi(b)d\rho = \int \sigma(a;1)\sigma(a;3)\Phi(b)d\rho. \tag{4.10}$$

From (4.9) and (4.10) we deduce the following equation for $g_1^{(2)}(a,b)$:

$$(1 - n_2 - n_4)[g_1^{(2)}(a,b)]^2 = A_3(1 - n_1 - n_3)g_1^{(2)}(a,b). \tag{4.11}$$

Here two cases are possible:

(α) $n_2 + n_4 \neq 1$ and $n_1 + n_3 \neq 1$. The two solutions of (4.11) are $g_1^{(2)}(a,b) = 0$, that is $\rho_1^{(2)}(a,b) = n_1^2$, which proves the lemma, or $g_1^{(2)}(a,b) = n_1(1 - n_1)$, that is, $\rho_1^{(2)}(a,b) = n_1$. This last solution cannot give a state satisfying condition (4.4). In fact $\rho_1^{(2)}(a,b) = n_1$ means that with a probability one the situation is the same on the lattice sites "a" and "b" when considering particles of velocity 1. From homogeneous property it results that with probability one the situation is the same on the family of sites $a + \tau_{r(b-a)}a$, where $r \in \mathbb{Z}$; therefore,

$$\int \sigma_X(a;1)\sigma_{\tau_{r(b-a)a}}(a;1)d\rho = n_1 \quad \forall\, r \in \mathbb{Z},$$

which contradicts (4.4).

(β) $n_2 + n_4 = 1$ and $n_1 + n_3 = 1$. In this case Eq. (4.11) is useless and we must construct another equation. It is easy to get from (3.1)

$$\sigma_{C(X)}(a;1)\sigma_{C(X)}(a;2) = \sigma_X(a;1)\sigma_X(a;2).$$

Thus C-invariance of ρ implies

$$\int \sigma(a;2)\sigma(a;1)\Phi(b)d\rho = 0. \tag{4.12}$$

Written in terms of correlation functions, (4.12) reads as

$$(n_2 - n_1)[g_1^{(2)}(a,b)]^2 - n_1(1 - n_1)(n_2 - n_1)g_1^{(2)}(a,b) = 0. \tag{4.13}$$

Now two cases are still possible.

(β') $n_1 \neq n_2$; We retrieve then the two solutions of case (α).

(β'') $n_1 = n_2$; and consequently $n_1 = n_2 = n_3 = n_4 = \frac{1}{2}$. Here we must use another equation. From (3.1) we get

$$\sigma_{C(X)}(a;1)\sigma_{C(X)}(a;2)\sigma_{C(X)}(a;3) = \sigma_X(a;1)\sigma_X(a;2)\sigma_X(a;3),$$

which leads to the following equation for the correlation functions:

$$[\tfrac{1}{4} + g_1^{(2)}(a,b)](g_1^{(2)}(a,b) - \tfrac{1}{4})g_1^{(2)}(a,b) = 0.$$

The new solution is $g_1^{(2)}(a,b) = -\frac{1}{4}$ which means that $\rho_1^{(2)}(a,b) = 0$. In other words with probability one the situation is different on the sites "a" and "b" when considering particles of velocity 1. From homogeneity assumption we conclude that with probability one the situation is the same on the family of sites $a + \tau_{2r(b-a)}a$, $\forall r \in \mathbb{Z}$, which contradicts (4.4). The proof of the lemma is thus achieved.

Now we can start with the induction argument.

Lemma 4.2: If $\rho \in \mathfrak{F}(n, v_x, v_y)$ is C-invariant and has no correlations up to order $2q$ with $q \geq 1$, then ρ has no correlations up to order $2q + 1$.

Proof: Consider $2q + 1$ distinct lattice sites $\{a_1, \ldots, a_{2q+1}\}$. We deduce from (3.1) that

$$\begin{aligned}\sigma_{C(X)}(a_1;1)\sigma_{C(X)}(a_2;1)\prod_{i=3}^{2q+1}\Phi_{C(X)}(a_i) &= -\sigma_X(a_1;1)\sigma_X(a_2;1)\prod_{i=3}^{2q+1}\Phi_X(a_i) \\ &+ \sigma_X(a_1;1)\Phi_X(a_2)\prod_{i=3}^{2q+1}\Phi_X(a_i) \\ &+ \sigma_X(a_2;1)\Phi_X(a_1)\prod_{i=3}^{2q+1}\Phi_X(a_i) - \prod_{i=1}^{2q+1}\Phi_X(a_i). \end{aligned} \tag{4.14}$$

Taking the average of (4.14) and using the assumptions, one gets

$$A_1^{2q-1}g_1^{(2q+1)}(a_1, \ldots, a_{2q+1}) = A_1^{2q}g_1^{(2q+1)}(a_1, \ldots, a_{2q+1}) \tag{4.15}$$

where $g_1^{(2q+1)} = \rho_1^{(2q+1)} - n_1^{2q+1}$. Using the fact that $0 < A_1 < 1$ (see Appendix A), we have proved the lemma.

Lemma 4.3: If $\rho \in \mathfrak{F}(n, v_x, v_y)$ is C-invariant and has no correlations up to order $2q + 1$ with $q \geq 1$, then ρ has no correlations up to order $2q + 2$.

Proof: Here we use the following equations:

$$\int \sigma_{C(X)}(a_1;1)\sigma_{C(X)}(a_2;1)\sigma_{C(X)}(a_3;1)\prod_{i=4}^{2q+2}\Phi_{C(X)}(a_i)d\rho = \int \sigma_X(a_1;1)\sigma_X(a_2;1)\sigma_X(a_3;1)\prod_{i=4}^{2q+2}\Phi_X(a_i)d\rho,$$

$$\int \sigma_{C(X)}(a_1;1)\prod_{i=4}^{2q+2}\Phi_{C(X)}(a_i)d\rho = \int \sigma_X(a_1;1)\prod_{i=2}^{2q+2}\Phi_X(a_i)d\rho.$$

Using the assumptions, we get finally

$$A_1^{2q-1}(A_1 - 1)(A_1 - 2)g_1^{(2q+2)} = 0.$$

As $0 < A_1 < 1$, the lemma is proved.

It is clear now that properties (c) and (d) characterize a unique state which we denote by $\rho(n, v_x, v_y)$. It is an elementary task to verify that this state is invariant under T and the proof of the theorem is finished. ■

Before going further, we make two remarks about

the assumptions of Theorem 4.1. First we want to show that condition (4.3) is not sufficient in order to ensure the uniqueness of a T-invariant state with given (n, v_x, v_y). In fact let us consider the state ρ' characterized by the following conditions: ρ' is homogeneous and factorized, $\rho' = \Pi_{i=1}^{4} \otimes \rho_i'$, ρ' has the same n_i as $\rho(n, v_x, v_y)$, there is no correlation between sites not lying on the same vertical, and finally for the site lying in the same vertical the correlations are such that with probability one the situation is the same on these sites, namely $\int \Pi_{i=1}^{r} \sigma_X(p, q_i; j) d\rho' = n_j \ \forall\, p \in \mathbb{Z}, \forall\, r \in \mathbb{N}$, $\forall\, j \in P$. The state ρ' is T-invariant. To prove this, it suffices to show that

$$\int \prod_{i=1}^{r} \sigma_{C(X)}(p, q_i; j) d\rho' = n_j \quad \forall\, p \in \mathbb{Z}, \forall\, r \in \mathbb{N}, \forall\, j \in P.$$

But it results from the assumptions that $\sigma_X(p, q_i j) = \sigma_X(p, q_i, j)$ a.e. with respect to ρ'. Then an elementary calculation shows that $\Pi_{i=1}^{r} \sigma_{C(X)}(p, q_i; j) = \sigma_{C(X)}(p, q_1; j)$ a.e., which proves the statement. Now it is clear that ρ' does not satisfy (4.4). To see that (4.3) is satisfied, we consider $A, B \in \mathfrak{A}$ and we may suppose that A and B depend only on a finite subset Λ_0 of lattice sites. Now there exists p_0 such that

$$\frac{1}{N(\Lambda)} \sum_{(p,q) \in \Lambda} \rho'(A \cap \tau_{(p,q)} B) = \frac{1}{N(\Lambda)} \sum_{\substack{|p| \le p_0 \\ (p,q) \in \Lambda}} \rho'(A \cap \tau_{(p,q)} B) + \frac{1}{N(\Lambda)} \sum_{\substack{|p| > p_0 \\ (p,q) \in \Lambda}} \rho'(A) \rho'(B).$$

Therefore

$$\lim_{\Lambda \to \infty} \frac{1}{N(\Lambda)} \sum_{(p,q) \in \Lambda} \rho'(A \cap \tau_{(p,q)} B) = \rho'(A) \rho'(B).$$

We have then two states $\rho(n, v_x, v_y)$ and ρ' having the same hydrodynamic parameters, invariant under T and satisfying (4.3).

In the assumptions of Theorem 4.1 we excluded the boundary of the domain Δ. Let us precise now what occurs on this boundary. We know that if (n, v_x, v_y) belongs to the boundary of Δ, one of the four numbers n_i is equal to one or zero. Therefore, there is a velocity "i" for which all the sites are occupied or not. Then, in order for a state $\rho \in \mathfrak{F}(n, v_x, v_y)$ to be C-invariant, it is necessary that with probability one no collision occurs, that is $C = 1$ a.e. This means that $T = T_0$ a.e. with respect to ρ. The time evolution is trivial in this case and this is why we disregarded the boundary of Δ.

Note that the state $\rho(n, v_x, v_y)$ defined by the Theorem 4.1 depends analytically on (n, v_x, v_y) in Δ in the sense that every expectation value of the form $\int \Pi_{p,q;i} \sigma_X(p, q; i) d\rho$ depends analytically on (n, v_x, v_y) in Δ. This is due to the fact that (see Appendix B) the solution χ of Eq. (4.7) is an analytic function of (n, v_x, v_y) in Δ and thus the quantities $n_i(n, v_x, v_y)$ also.

We conclude this section by showing that the state $\rho(n, v_x, v_y)$ satisfies a variational principle. More precisely let $\rho \in \mathfrak{F}(n, v_x, v_y)$. We can define its entropy by (3.3). As $\rho = \Pi_{i=1}^{4} \otimes \rho_i$, we have in fact $H(\rho) = \frac{1}{4} \sum_{i=1}^{4} H(\rho_i)$, where

$$H(\rho_i) = \lim_{\Lambda \to \infty} \frac{1}{N(\Lambda)} \sum_{X_i \subset \Lambda} [-\rho_{i,\Lambda}(X_i) \log \rho_{i,\Lambda}(X_i)].$$

Proposition 4.3: Let $\rho \in \mathfrak{F}(n, v_x, v_y)$ then ρ is T-invariant if and only if its entropy $H(\rho)$ is maximum.

Proof: It suffices to show that $H(\rho)$ reaches its maximum in a unique point which is precisely the state $\rho(n, v_x, v_y)$. Using the fact that $H(\rho) = \frac{1}{4} \sum_{i=1}^{4} H(\rho_i)$, we can assert that given (n_1, n_2, n_3, n_4) the maximum of $H(\rho)$ is reached iff each $H(\rho_i)$ is maximum. But the maximum of $H(\rho_i)$ when n_i is given is reached in a unique point which is the state without correlation; in addition,

$$\max H(\rho_i) = -[n_i \log n_i + (1 - n_i) \log(1 - n_i)].$$

Now is it an elementary calculation to show that $-\frac{1}{4} \sum_{i=1}^{4} [n_i \log n_i + (1 - n_i) \log(1 - n_i)]$ is maximum in the domain $\sum_{i=1}^{4} n_i = n$, $n_1 - n_3 = v_x$, $n_2 - n_4 = v_y$ if and only if (4.5) holds. The proposition is then proved. In this section homogeneous states were only considered. For completeness let us mention that it is easy to construct nonhomogeneous T-invariant states. In fact consider a state ρ factorized. We set $n_i(p, q) = \int \sigma_X(p, q; i) d\rho$. Let us choose $n_1(p, q)$ and $n_3(p, q)$ depending only on q, in the same way let $n_2(p, q)$ and $n_4(p, q)$ be depending only on p in such a way that

$$n_1(q) n_3(q) / [1 - n_1(q)][1 - n_3(q)] = n_2(p) n_4(p) / [1 - n_2(p)][1 - n_4(p)] = cte.$$

It is clear that such a state is nonhomogeneous. But this state is T-invariant because it is separately T_0-invariant and C-invariant.

5. ERGOTIC PROPERTIES AND TIME CORRELATION FUNCTIONS

In the remainder of this paper we shall fix our attention on the state $\rho(n, v_x, v_y)$ and investigate its ergodic properties with respect to T. For brievity we shall omit the parameters (n, v_x, v_y).

Consider the dynamical system $\{K, T, \rho\}$, we recall that this system is said to be ergodic[11] if

$$\lim_{t \to \infty} \frac{1}{t} \sum_{\tau=0}^{t-1} \rho(A \cap T^{\tau} B) = \rho(A) \rho(B) \quad \forall\, A, B \in \mathfrak{A}. \tag{5.1}$$

The system is said to be mixing[11] if

$$\lim_{\tau \to \infty} \rho(A \cap T^{\tau} B) = \rho(A) \rho(B) \quad \forall\, A, B \in \mathfrak{A}. \tag{5.2}$$

Definitions (5.1) and (5.2) are particular cases of (4.3) and (4.4). It is obvious that (5.2) implies (5.1). We can rewrite the condition (5.2) in an equivalent form:

$$\lim_{t \to \infty} \int f(X) g(T^t X) d\rho = \int f d\rho \int g d\rho \quad \forall\, f, g \in L^2(K, \rho). \tag{5.3}$$

Let us show that (5.3) is necessary and sufficient in order to ensure the approach towards equilibrium of local perturbations of the system. Consider a local observable that is a function $\varphi(X) : K \to \mathbb{R}$ such that $\varphi(X)$ depends only on the part of X lying in a finite subset Λ of $\mathbb{Z}^2$. Such a function takes a finite number of different values and belongs to $\mathbb{C}(K)$, the space of complex continuous function defined on K. Note that any $f \in L^2(K, \rho)$ can be approximated by a local observable in the norm of $L^2(K, \rho)$, so that (5.3) is equivalent to

$$\lim_{t \to \infty} \int f(X) \varphi(T^t X) d\rho = \int f d\rho \int \varphi \rho \quad \forall\, f \in L^2(K, \rho) \quad \forall\, \varphi \text{ a local observable.} \tag{5.3'}$$

We can consider φ as a perturbation to the initial Hamiltonian of the system. The thermodynamic equilibrium state ρ' corresponding to the total Hamiltonian can be written as:

$$\rho' = \{\exp(-\beta\varphi)/[\int \exp(-\beta\varphi)d\rho]\}\rho, \tag{5.4}$$

where ρ means $\rho(n, v_x, v_y)$. Such a state is called a local perturbation of the equilibrium state ρ. We have then

Proposition 5.1: Any local perturbation of the equilibrium state ρ relaxes as $t \to \infty$ towards ρ if and only if ρ satisfies the mixing property (5.3).

Proof: Let ρ' be a state given by (5.4). We can write $\rho' = \psi(X)\rho$ where $\psi(X)$ is a positive local observable such that $\int \psi d\rho = 1$. Now (5.3′) implies that ρ' relaxes towards ρ in the sense of the vague topology[12] on the set of states on K. Conversely if (5.3′) holds whenever φ is a positive local observable such that $\int \varphi d\rho = 1$ by linearity it holds for any local observable and the proposition is proved.

Before investigating the mixing properties of the system $\{K, T, \rho\}$ let us make the following simple remarks on the much simpler system $\{K, T_0, \rho\}$, where T_0 is the free evolution. In fact we have

Proposition 5.2: The system $\{K, T_0, \rho\}$ is a Bernouilli shift[11] of infinite entropy.

Proof: For completeness we recall that a Bernouilli shift of infinite entropy is a dynamical system (K, T, ρ) such that there exists an uncountable partition P of K such that $K = \Pi_{-\infty}^{+\infty} T^n P$ and ρ factorizes under that decomposition.

In order to prove the proposition, let us call τ_i the unit translation in the direction of velocity "i", $i \in P$. Then $\{K, T, \rho\} = \Pi_{i=1}^{4}\{K_i, \tau_i, \rho_i\}$. Now $\{K_i, \tau_i, \rho_i\}$ is a Bernouilli shift of infinite entropy.

It is known[13] that Bernouilli shifts have the strongest mixing properties among all dynamical systems. Therefore, (5.3) is satisfied, if T is replaced by T_0. Now if (5.3) is satisfied for the free evolution it is reasonable to expect that (5.3) is still true for the evolution with collisions. Actually the presence of collisions should increase the mixing in the velocity space. In order to evaluate the efficiency of the collisions, let us calculate for instance the exact collision frequency for a particle in the system.

Proposition 5.3: The mean free time between two successive collisions for a particle of velocity i, is given by

$$t_i = 1/n_{i+2}(1 - n_{i+1})(1 - n_{i+3}), \tag{5.5}$$

where the n_i are given by (4.2) and satisfy (4.5).

Proof: Let us prove (5.5) for $i = 1$ for instance. We introduce the following characteristic function:

$$A_t(X, 1) = \bar{\sigma}_{T^{-1}(X)}(-1, 0; 1)\left(\prod_{j=0}^{t-1}\sigma_{T^j(X)}(j, 0; 1)\right)\bar{\sigma}_{T^t(X)}(t, 0; 1)$$

$$= \begin{cases} 1 & \text{if a particle of velocity 1 is created at the origin at time 0, moves freely during the time interval } [0, t-1] \text{ and suffers a collision at time } t, \\ 0 & \text{otherwise.} \end{cases}$$

We have $t_1 = \sum_{t=0}^{\infty} t\int A_t(X, 1)d\rho / \int \bar{\sigma}_{T^{-1}(X)}(-1, 0; 1)\sigma_X(0, 0, 1)d\rho$. Now we remark that T-invariance of ρ implies that

$$\int A_t(X, 1)d\rho = \int \bar{\sigma}_{T^{-t}(X)}(-1, 0; 1)\left(\prod_{j=0}^{t-2}\sigma_{T^{j-t+1}(X)}(j, 0; 1)\right) \times \sigma_X(t-1, 0; 1)\bar{\sigma}_{T(X)}(t, 0; 1)d\rho.$$

It follows from (2.1) that

$$\sigma_X(t-1, 0; 1)\bar{\sigma}_{T(X)}(t, 0, 1) = \sigma_X(t-1, 0; 1)\sigma_X(t+1, 0; 3) \times \bar{\sigma}_X(t, -1; 2)\bar{\sigma}_X(t, 1; 4)$$

and that

$$\bar{\sigma}_{T^{-t}(X)}(-1, 0; 1)\left(\prod_{j=0}^{t-2}\sigma_{T^{j-t+1}(X)}(j, 0; 1)\right)\sigma_X(t-1, 0; 1)$$

is independent of $\sigma_X(t+1, 0; 3)\bar{\sigma}_X(t, -1; 2)\bar{\sigma}_X(t, 1; 4)$, so that we have

$$\int A_t(X, 1)d\rho = n_3(1-n_2)(1-n_4)\int \bar{\sigma}_{T^{-1}(X)}(-1, 0; 1) \times \left(\prod_{j=0}^{t-2}\sigma_{T^j(X)}(j, 0; 1)\right)\sigma_{T^{t-1}(X)}(t-1, 0; 1)d\rho.$$

By induction one gets easily the following equation:

$$\int A_t(X, 1)d\rho = n_3(1-n_2)(1-n_4)[1-n_3(1-n_2)(1-n_4)]^{t-1} \times \int \bar{\sigma}_{T^{-1}(X)}(-1, 0; 1)\sigma_X(0, 0; 1)d\rho;$$

therefore,

$$t_1 = \sum_{t=1}^{\infty} t\, n_3(1-n_2)(1-n_4)[1-n_3(1-n_2)(1-n_4)]^{t-1} = 1/n_3(1-n_2)(1-n_4). \quad \blacksquare$$

Note that t_i is also the mean free path for a particle of velocity i because each particle has a velocity equal to one in absolute value.

We return now to the investigation of (5.3). We remark that it suffices to construct a complete orthogonal system of functions for which (5.3) holds. In our case we can take the following basis. Let us set $s_X(p, q; i) = \sigma_X(p, q; i) - n_i$, we denote by $\Pi_{p,q;i} s_X(p, q; i)$ any finite product of different functions $s_X(p, q; i)$. This set of orthogonal functions generates the subspace of functions of $L^2(K, \rho)$ such that $\int f d\rho = 0$; therefore, (5.3) is equivalent to

$$\lim_{t\to\infty} \int \prod_{p,q;i} s_{T^t(X)}(p, q; i) \prod_{p',q';i'} s_X(p', q'; i')d\rho = 0. \tag{5.6}$$

The following program could then be adopted: Write Eqs. (2.1) in terms of the functions $s_X(p, q; i)$, use the fact that $s_X^2(p, q; i) = (1 - 2n_i)s_X(p, q; i) + n_i(1 - n_i)$ to write $\Pi_{p,q;i} s_{T^t(X)}(p, q; i)$ as a polynomial of the first degree with respect to each function $s_X(p, q; i)$; from the orthogonality property it results that the coefficient of the monomial $\Pi_{p',q';i'} s_X(p', q'; i')$ is exactly the time correlation function in the left-hand side of (5.6). The problem is thus to prove that the coefficients of this polynomial vanish asymptotically when the process is iterated.

In this paper we do not pretend to achieve such a program. First of all we shall restrict our considerations to the simplest time correlation functions $\int s_{T^t(X)}(p, q; i)\, s_X(p', q'; i')$ in the particular case where $v_x = v_y = 0$. In addition we shall make an approximation on the evolution equations for $s_X(p, q; i)$ which will allow us to calculate the long time tail of these correlation functions.

6. THE LINEAR APPROXIMATION

Let us write the Eqs. (2.1) in terms of the functions $s_X(p,q;i)$ in the case where $v_x = v_y = 0$, that is, $n_1 = n_2 = n_3 = n_4 = n_0$,

$$\begin{aligned} s_{T(X)}(p,q;1) &= s_X(p-1,q;1) - \Theta_X(p,q), \\ s_{T(X)}(p,q;2) &= s_X(p,q-1;2) + \Theta_X(p,q), \\ s_{T(X)}(p,q;3) &= s_X(p+1,q;3) - \Theta_X(p,q), \\ s_{T(X)}(p,q;4) &= s_X(p,q+1;4) + \Theta_X(p,q), \end{aligned} \tag{6.1}$$

where

$$\begin{aligned} \Theta_X(p,q) = {} & \mu s_X(p-1,q;1) - \mu s_X(p,q-1;2) \\ & + \mu s_X(p+1,q;3) - \mu s_X(p,q+1;4) \\ & + (1-2n_0)[s_X(p,q-1;2)s_X(p,q+1;4) \\ & - s_X(p-1,q;1)s_X(p+1,q;3)] \\ & + s_X(p,q-1;2)s_X(p,q+1;4) \\ & \times [s_X(p-1,q;1) + s_X(p+1,q;3)] \\ & - s_X(p-1,q;1)s_X(p+1,q;3) \\ & \times [s_X(p,q-1;2) + s_X(p,q+1;4)] \end{aligned}$$

with $\mu = n_0(1-n_0)$.

Our approximation consists in neglecting in Eqs. (6.1) the nonlinear terms with respect to the functions $s_X(p,q;i)$. We get then

$$\begin{aligned} s_{T(X)}(p,q;1) &= (1-\mu)s_X(p-1,q;1) + \mu s_X(p,q-1;2) \\ &\quad - \mu s_X(p+1,q;3) + \mu s_X(p,q+1;4), \\ s_{T(X)}(p,q;2) &= \mu s_X(p-1,q;1) + (1-\mu)s_X(p,q-1;2) \\ &\quad + \mu s_X(p+1,q;3) - \mu s_X(p,q+1;4), \\ s_{T(X)}(p,q;3) &= -\mu s_X(p-1,q;1) + \mu s_X(p,q-1;2) \\ &\quad + (1-\mu)s_X(p+1,q;3) + \mu s_X(p,q+1;4), \\ s_{T(X)}(p,q;4) &= \mu s_X(p-1,q;1) - \mu s_X(p,q-1;2) \\ &\quad + \mu s_X(p+1,q;3) + (1-\mu)s_X(p,q+1;4). \end{aligned} \tag{6.2}$$

It is natural to introduce the Fourier transform of $s_X(p,q;i)$:

$$\tilde{S}(t,z_1,z_2;i) = \sum_{p=-\infty}^{+\infty} \sum_{q=-\infty}^{+\infty} z_1^p z_2^q s_{T^t(X)}(p,q;i). \tag{6.3}$$

After substituting (6.3) into (6.2) we get

$$\tilde{S}(t,z_1,z_2) = M(z_1,z_2)\tilde{S}(t-1,z_1,z_2),$$

where

$$\tilde{S}(t,z_1,z_2) = \begin{vmatrix} \tilde{s}(t,z_1,z_2;1) \\ \tilde{s}(t,z_1,z_2;2) \\ \tilde{s}(t,z_1,z_2;3) \\ \tilde{s}(t,z_1,z_2;4) \end{vmatrix} \quad \text{and}$$

$$M(z_1,z_2) = \begin{vmatrix} (1-\mu)z_1 & \mu z_2 - \mu\bar{z}_1 & \mu\bar{z}_2 \\ \mu z_1 & (1-\mu)z_2 & \mu\bar{z}_1 - \mu\bar{z}_2 \\ -\mu z_1 & \mu z_2 & (1-\mu)z_1\mu\bar{z}_2 \\ \mu z_1 & -\mu z_2 & \mu\bar{z}_1 \quad (1-\mu)\bar{z}_2 \end{vmatrix} \tag{6.4}$$

By Fourier inversion we have

$$S(t,p,q) = (2i\pi)^{-2}\int_{C_1}\int_{C_2} M^t(z_1,z_2)\tilde{S}(0,z_1,z_2)\frac{dz_1 dz_2}{z_1^{p+1}z_2^{q+1}} \tag{6.5}$$

Here C_1 (resp. C_2) means the unit circle in the complex z_1 (resp. z_2) plane.

We are interested in the asymptotic part of the right-hand side of (6.5) as $t \to \infty$. To evaluate $M^t(z_1,z_2)$, we introduce the eigenvalues $\{\lambda_\alpha\}$, $\alpha \in (1,2,3,4)$, of $M(z_1,z_2)$ and the corresponding eigenvectors $\{\epsilon_\alpha\}$ that we suppose to be distinct for the moment. Let us call $\{e_i\}$ the natural basis of R^4, we set $\delta_i^\alpha = (\epsilon_\alpha \cdot e_i)$ where $(\cdot)$ means the Hermitian product and we set δ_α^i, $\alpha \in (1,2,3,4)$, the components of e_i in the basis $\{\epsilon_\alpha\}$. We must be careful that $\delta_\alpha^i \neq \delta_i^\alpha$ because $\{\epsilon_\alpha\}$ is not in general an othogonal basis. We finally have

$$\begin{aligned} \int s_X(0,0;i)s_{T^t(X)}(p,q;j)d\rho &= \mathcal{C}(p,q;i,j;t) \\ &= \frac{\mu}{(2i\pi)^2}\sum_{\alpha=1}^{4}\int_{C_1\times C_2}\frac{dz_1 dz_2}{z_1^{p+1}z_2^{q+1}}(\lambda_\alpha)^t\delta_i^\alpha\delta_\alpha^j. \end{aligned} \tag{6.6}$$

The following proposition gives some qualitative results about the eigenvalues $\{\lambda_\alpha\}$ which will permit us to reduce the integration domain in order to evaluate the asymptotic part as $t \to \infty$ of the right-hand side of (6.6).

Proposition 6.1: (a) $|\lambda_\alpha| \leq 1 \; \forall \; \alpha \in (1,2,3,4)$ $\forall \; z_1, z_2 \in C_1 \times C_2$.

(b) The regions of $C_1 \times C_2$ where there exists at least one eigenvalue of modulus one, are the following:

$$\mathcal{R}_1: \begin{cases} z_1 = +1 \\ z_2 = +1 \end{cases} \left(\text{resp.} \begin{matrix} z_1 = -1 \\ z_2 = -1 \end{matrix}\right),$$

$$\lambda_1 = \lambda_2 = \lambda_3 = 1 \; (\text{resp.} -1)$$

with the corresponding eigenvectors

$$\begin{vmatrix} 1 \\ 1 \\ 1 \\ 1 \end{vmatrix}, \begin{vmatrix} 1 \\ 0 \\ -1 \\ 0 \end{vmatrix}, \begin{vmatrix} 0 \\ 1 \\ 0 \\ -1 \end{vmatrix};$$

$$\mathcal{R}_2 = \begin{cases} z_1 = 1 \\ z_2 = -1 \end{cases} \left(\text{resp.} \begin{matrix} z_1 = -1 \\ z_2 = 1 \end{matrix}\right), \quad \lambda_1 = -\lambda_2 = 1$$

with the corresponding eigenvectors

$$\begin{vmatrix} 0 \\ 1 \\ 0 \\ -1 \end{vmatrix}, \begin{vmatrix} 1 \\ 0 \\ -1 \\ 0 \end{vmatrix};$$

$$\mathcal{R}_3: \begin{cases} z_1 = \pm 1 \\ z_2 \neq \pm 1 \end{cases} \left(\text{resp.} \begin{matrix} z_1 \neq \pm 1 \\ z_2 = \pm 1 \end{matrix}\right), \quad \lambda_1 = \pm 1$$

with the eigenvector

$$\begin{vmatrix} 1 \\ 0 \\ -1 \\ 0 \end{vmatrix} \left(\text{resp.} \begin{vmatrix} 0 \\ 1 \\ 0 \\ -1 \end{vmatrix}\right)$$

$$\mathcal{R}_4: \begin{cases} z_1 = z_2 \\ z_2 \neq \pm 1 \end{cases} \left(\text{resp.} \begin{matrix} z_1 = \bar{z}_2 \\ z_2 \neq \pm 1 \end{matrix}\right), \quad \lambda_1 = \bar{\lambda}_2 = z_1$$

with the eigenvectors

$$\begin{vmatrix} 1 \\ 1 \\ 0 \\ 0 \end{vmatrix}, \begin{vmatrix} 0 \\ 0 \\ 1 \\ 1 \end{vmatrix} \quad \left(\text{resp.} \begin{vmatrix} 0 \\ 1 \\ 1 \\ 0 \end{vmatrix} \; \begin{vmatrix} 1 \\ 0 \\ 0 \\ 1 \end{vmatrix}\right).$$

Proof: Note that $M(z_1, z_2) = C \cdot T_0$ with

$$C = \begin{vmatrix} 1-\mu & \mu & -\mu & \mu \\ \mu & 1-\mu & \mu & -\mu \\ -\mu & \mu & 1-\mu & \mu \\ \mu & -\mu & \mu & 1-\mu \end{vmatrix} \quad \text{and}$$

$$T_0 = \begin{vmatrix} z_1 & 0 & & \\ 0 & z_2 & & \\ & & \bar{z}_1 & \\ & & & \bar{z}_2 \end{vmatrix}$$

T_0 is a unitary matrix and C is symmetric, its eigenvalues are 1 and $1 - 4\mu < 1$, the corresponding eigenvectors are

$$\begin{vmatrix} 1 \\ 1 \\ 1 \\ 1 \end{vmatrix}, \quad \begin{vmatrix} 1 \\ 0 \\ -1 \\ 0 \end{vmatrix}, \quad \begin{vmatrix} 0 \\ 1 \\ 0 \\ -1 \end{vmatrix}, \quad \text{and} \quad \begin{vmatrix} 1 \\ -1 \\ 1 \\ -1 \end{vmatrix}.$$

Therefore the norm of a vector cannot be increased by the matrix $M(z_1, z_2)$ and this proves the part (a) of the proposition.

To prove (b) we remark that an eigenvalue of $M(z_1, z_2)$ is of modulus one if and only if there exists an eigenvector of T_0, which belongs to the invariant subspace with respect to the matrix C. This leads to the set of equations

$$\begin{aligned} z_1(\alpha+\beta) &= \lambda(\alpha+\beta), \quad & z_2(\alpha+\gamma) &= \lambda(\alpha+\gamma), \\ \bar{z}_1(\alpha-\beta) &= \lambda(\alpha-\beta), \quad & \bar{z}_2(\alpha-\gamma) &= \lambda(\alpha-\gamma). \end{aligned} \tag{6.7}$$

Now it is an elementary calculation to show that the only solutions of (6.7) such that $|\alpha|^2 + |\beta|^2 + |\gamma|^2 + |\delta|^2 \neq 0$ are those given in the statement of the proposition.

The above result deserves some remarks. First of all there is an obvious symmetry of the domains with respect to the mapping $(z_1, z_2) \leftrightarrow (-z_1, -z_2)$; this will permit us to reduce the region of integration in the right-hand side of (6.6). Secondly it is natural to consider the quantities $\sqrt{-1}\log\lambda_\alpha$, $\alpha \in (1, 2, 3, 4)$ as the frequencies of "modes" that can propagate in the system. The hydrodynamic regime corresponds to the region $z_1 \simeq 1$, $z_2 \simeq 1$. In realistic systems it is conjectured that this is the single region where the imaginary part of the frequencies, that is the damping rate, vanishes.[7] This is not the case in our system. Indeed in the region $\mathcal{R}_3$ there is a mode whose frequency vanishes. This means that when the system is driven by a perturbation homogeneous in the $0x$ (resp. $0y$) direction, there is a component of the perturbation which remains constant. This component is precisely the hydrodynamic momentum along the $0x$ (resp. $0y$) direction. This very peculiar property is a consequence of the fact that the impact parameter in a collision is always zero. There is another region of $C_1 \times C_2$, where some eigenfrequencies have a vanishing imaginary part. This is the region $\mathcal{R}_4$ where there are two real opposite eigenfrequencies. This can be interpreted by saying that when the system is driven by a perturbation homogeneous in the direction of one of the bissectrices of $(0x, 0y)$, there are two components of the perturbation which propagate without damping. This is a new consequence of the fact that the impact parameter is always zero.

In view of Proposition 6.1, in order to evaluate the asymptotic part as $t \to \infty$ of the right-hand side of (6.6), we can limit the integration in a small neighborhood of the region $\cup_{i=1}^4 \mathcal{R}_i$, the remainder of $C_1 \times C_2$ will give an exponentially decreasing contribution. We can then write

$$\mathcal{C}(p, q; i, j; t) \simeq \sum_{i=1}^{4} (\mathcal{R}_i) \quad \text{as } t \to \infty \tag{6.8}$$

with

$$(\mathcal{R}_i) = \frac{\mu}{(2i\pi)^2} \sum_{\alpha=1}^{4} \iint_{(z_1, z_2) \simeq \mathcal{R}_i} \frac{dz_1 dz_2}{z_1^{p+1} z_2^{q+1}} \lambda_\alpha^t \delta_i^\alpha \delta_\alpha^j.$$

The calculation of each contribution is rather messy and is reported in Appendix C. The result is the following:

$$\mathcal{C}(p, q; i, j, t) \simeq c(i, j) t^{-1/2} \quad \begin{cases} \text{if } i + j \text{ even and } i \text{ odd,} \\ \text{with } q = 0 \text{ or } i \text{ even,} \\ \text{with } p = 0, \end{cases} \tag{6.9a}$$

$$\mathcal{C}(p, q; i, j, t) \simeq c(p, q; i, j) t^{-3/2} \quad \text{in the other cases.} \tag{6.9b}$$

The values for the coefficients $c(i, j)$ and $c(p, q; i, j)$ are given in Table CH of Appendix C. We should emphasize the fact that the estimate (6.9b) is valid unless the coefficient $c(p, q; i, j)$ is not finite. Unfortunately we have no satisfactory proof of this property at present.

7. DISCUSSION AND CONCLUSION

We give in this section some comments on the above results. First of all one remarks that the long time tails given in (6.9) are rather unexpected. In fact, for two-dimensional systems, arguments based on the theory of propagation of long wave hydrodynamical modes[7,14] predict long time tails like t^{-1} for the time correlation function of the momentum at two different points in the fluid. In our model we recovered the influence of some long wave modes, and it is noticed in Appendix C that only one mode contributes to the long time tail. This mode is called the vorticity diffusion mode by analogy with the other systems, but, as we shall see below there is, strictly speaking, no vorticity diffusion in this model and this is why we have long time tails like $t^{-1/2}$ or $t^{-3/2}$. In fact let us introduce the microscopic momentum at the lattice point (p, q):

$$v_X(p, q; x) = s_X(p, q, 1) - s_X(p, q; 3)$$

and

$$v_X(p, q; y) = s_X(p, q; 2) - s_X(p, q; 4).$$

We recall the fact, mentioned in Sec. 2, that the horizontal (resp. vertical) momentum is conserved on each horizontal (resp. vertical) line of lattice sites. This suggests that the horizontal (resp. vertical) momentum at the origin diffuses on the x (resp. y) axis like a one-dimensional diffusion process. Therefore, one should have

$$\int v_X(0, 0; x) v_{T^t(X)}(p, 0; x) d\rho \simeq c t^{-1/2},$$

$$\int v_X(0, 0; y) v_{T^t(X)}(0, q; y) d\rho \simeq c t^{-1/2}.$$

This is exactly in agreement with (6.9a). For the other correlation functions estimated in (6.9b) we have at present no satisfactory explanation of the long time tail like $t^{-3/2}$. Nevertheless, it is clear that this is again reminiscent of the fact that the impact parameter is zero for all collisions.

The results (6.9) have been obtained after a linearization of Eq. (6.1). Let us make some remarks concerning the meaning of this linear approximation. We

would like to show how this approximation can be connected with the linearized Boltzmann equation. To see this, consider a nonhomogeneous probability measure ρ on K which is not T-invariant. We set $n(p,q;i) = \int \sigma_X(p,q;i)d\rho$. The four numbers $n(p,q;i)$, $i \in P$, are the analog of the velocity distribution function at the point (p,q) in a classical fluid. From the four numbers $n(p,q;i)$ we can calculate the three hydrodynamic parameters:

$$n(p,q) = \sum_{i=1}^{4} n(p,q;i), \qquad v_x(p,q) = \frac{n(p,q;1) - n(p,q;3)}{n(p,q)},$$

$$v_y(p,q) = \frac{n(p,q;2) - n(p,q;4)}{n(p,q)}.$$

Let us suppose for simplication that these parameters are independent of (p,q) and that $v_x = v_y = 0$, $n = 4n_0$. We set $n(p,q;i) = n_0 + \delta n(p,q;i)$, where $\delta n(p,q;i)$ is the deviation from the local equilibrium. One can take the average with respect to ρ of Eqs. (2.1), one gets then a set of equations which are the analogs of the first equation of the BBGKY hierarchy. If one neglects the initial correlations (stosszahlansatz) and the nonlinear terms in $\delta n(p,q;i)$, one has exactly Eqs. (6.2), where $s_X(p,q;i)$ is replaced by $\delta n(p,q;i)$, and the approximations we have made to arrive at these equations are very similar to the assumption for deriving the linearized Boltzmann equation from the BBGKY hierarchy. This suggests that our linear approximation is probably valid in the case of small values of μ, which corresponds here to the case of low density $n_0 \simeq 0$ as well as high density $n_0 \simeq 1$. In view of the above remarks, the existence of modes very similar to the hydrodynamical modes when computing the asymptotic part of time correlation function is not surprising. The interesting point is that we started in Sec. 6 from microscopic equations and not from kinetic equations.

There is another interesting property of the linear approximation which must be pointed out. In fact the linear approximation gives the exact values for the correlation functions like

$$\int s_{T^t(X)}(0,0;i)s_X(p,q;j)d\rho \quad \text{with } |p| + |q| = t,$$

that is, we are able to calculate exactly the time correlation between the initial perturbation at the origin and the resulting perturbation at time t on the boundary of the perturbed region. To see this, we recall that $s_{T^t(X)}(0,0;i)$ can be written as a polynomial with respect to the functions $s_X(p,q;j)$. We can make the two following remarks about this polynomial:

(a) In the polynomial at time t there is no $s_X(p,q;j)$ with $|p| + |q| > t$. In addition there is no $s_X(p,q;j)$ with $|p| + |q| = t$ such that j corresponds to a velocity directed outside the square $|p| + |q| \leq t$.

(b) When iterating from time t to time $t+1$, only terms like $s_X(p,q;i)s_X(p,q;j)$ with $j = i + 2$ and like $s_X(p,q;j)s_X(p,q;i)s_X(p,q;k)$ with $i \neq j \neq k$ in the polynomial at time t can contribute to the linear terms of the polynomial at time $t+1$.

These two remarks are direct consequences of Eqs. (6.1) and the fact that $s_X^2(p,q;i) = (1 - 2n_0)s_X(p,q;i) + \mu$. In view of the above remarks we immediately see that the linear terms in $s_{T^{t+1}(X)}(0,0;j)$ with $|p| + |q| = t+1$ come only from the linear terms of $s_{T^t(X)}(0,0;j)$ with $|p| + |q| = t$. This allows us to calculate $\int s_{T^t(X)}(0,0;j)s_X(p,q;i)d\rho$ when $|p| + |q| = t$.

In Appendix D we compute the long time tail of this last quantity, in the two characteristic cases, p fixed and $|p| - |q|$ fixed. The results are

$$\mathcal{C}(p,q;i,j,t) \simeq c \times (1-\mu)^{t|p|} \quad \text{when } |p| + |q| = t,\ p \text{ fixed} \qquad (7.1a)$$

$$\mathcal{C}(p,q;i,j,t) \simeq c \times t^{-1/2} \text{ or } c \times t^{-3/2} \quad \text{when } |p| + |q| = t,\ |p| - |q| \text{ fixed} \qquad (7.1b)$$

In formula (7.1b) the two estimates $c \times t^{-1/2}$ and $c \times t^{-3/2}$ depend on the respective values of i and j. The cases are made precise in Appendix D. The interesting point in formula (7.1) is that the long time tail is quite different according as p (resp. q) is fixed or $|p| - |q|$ is fixed. The meaning of the above results is the following; if the system is perturbed at the origin at time zero, then the resulting perturbation at time t, at the point $(0,t)$ for instance decreases exponentially at $t \to \infty$ while the resulting perturbation at time t at the point $(t/2, t/2)$ for instance, decreases as $t^{-1/2}$ or $t^{-3/2}$. In other words there is a part of the initial perturbation which travels with a velocity $1/\sqrt{2}$ and which is slowly damped. It is natural to think that this is the part of the perturbation which is carried by the sound waves. Indeed $1/\sqrt{2}$ is exactly the group velocity of the modes called sound modes in the Table C1 of Appendix C, because they were the only modes which could propagate in the hydrodynamic regime. In some sense the above result is a rigorous proof of the existence of sound waves of velocity $1/\sqrt{2}$ in this system.

Let us mention another interesting property of Eq. (6.1). In fact in the case $n_0 = \frac{1}{2}$ there is a simplification in these equations, namely the quadratic terms disappear. This implies that $s_{T^t(X)}(0,0;i)$ is a polynomial of odd degree with respect to $s_X(p',q';i')$. This is reminiscent of the fact that when $n_0 = \frac{1}{2}$ the state $\rho(n_0)$ is invariant under the mapping $\mathcal{L}$ (see Sec. 2) and that $s_{(X)}(p,q;i) = -s_X(p,q;i)$ if $n_0 = \frac{1}{2}$, so that

$$\int \Pi s_{T^t(X)}(p,q;i) \Pi s_X(p',q';i')d\rho = 0$$

whenever the two products are of different parities. Eventually this simplification can lead to rigorous results in the evaluation of time correlation functions. This point is currently under investigation.

We shall report in a subsequent paper the results obtained on the hydrodynamic description of the model and the definition of some sort of viscosity coefficient.

ACKNOWLEDGMENTS

The authors are gratefully indebted to D. Ruelle, E. Hauge, E. G. D. Cohen, and E. Lieb for discussions and to D. Robinson for reading the manuscript. One of use (O. de Pazzis) is grateful to Professor A. Visconti for his kind hospitality at the Centre de Physique Théorique de Marseille.

APPENDIX A

Lemma: If $(n, v_x, v_y) \in \Delta$ and if (4.5) holds, then $0 < A_i < 1$.

Proof: First of all note that if $(n, v_x, v_y) \in \Delta$ and if (4.5) holds, then $0 < n_i < 1$. It results that $A_i > 0$.

On the other hand A_i is a continuous function of (n, v_x, v_y) on the convex Δ because the solution χ of

Eq. (4.7) is analytic in (n, v_x, v_y) (see Appendix B). The image of Δ under A_i is thus an interval of R^+ containing the number $\frac{1}{4}$ which is the particular value of A_i for $v_x = v_y = 0$, $n = 2$. It suffices then to show that $A_i = 1$ is impossible. Let us consider, for instance, the equation $A_1 = 1$ together with (4.5):

$$n_1 = n_2 n_4 (1 - n_3), \qquad 1 - n_1 = n_3 (1 - n_2)(1 - n_4)$$

so that $n_3 = (1 - n_2 n_4)/(1 - n_2 - n_4)$. But if $0 < n_2 < 1$ and $0 < n_1 < 1$, then either $1 - n_2 - n_4 \leq 0$ which implies $n_3 \leq 0$, or $1 - n_2 - n_4 \geq 0$ which implies $n_3 \geq 1$. The two conclusions are impossible and the lemma is proved.

APPENDIX B

The solution χ of (4.7) which satisfies conditions (4.2) depends analytically on (n, v_x, v_y) in Δ. Note that (4.7) reduces to

$$\chi^3 + \tfrac{1}{4}\{n(1 - \tfrac{1}{4}n) - \tfrac{1}{2}[(nv_x)^2 + (nv_y)^2]\}\chi + \tfrac{1}{16}(1 - \tfrac{1}{2}n)[(nv_y)^2 - (nv_x)^2] = 0.$$

This equation is of the form $\chi^3 + p\chi + q = 0$ with

$$p = \tfrac{1}{4}\{n(1 - n/4) - \tfrac{1}{2}[(nv_x)^2 + (nv_y)^2]\},$$

$$q = \tfrac{1}{16}(1 - \tfrac{1}{2}n)[(nv_y)^2 - (nv_x)^2].$$

We remark that if $(n, v_x, v_y) \in \Delta$ then

$$(nv_x)^2 + (nv_y)^2 \leq |nv_x| + |nv_y| \leq \min(n, 4 - n) \leq n(4 - n)/2$$

so that $p \geq 0$. Therefore, in order that $4p^3 + 27q^2 = 0$, we must have $p = q = 0$. It is easy to see that the only possibility in $\bar{\Delta}$ is $n = 2$, $|nv_x| = |nv_y| = 1$ but these points belong to the boundary of Δ. This shows that the solution of (4.7) is an analytic function of (n, v_x, v_y) in Δ, and therefore the same property holds for n_i, $i \in P$.

APPENDIX C

In order to evaluate (6.6) as $t \to \infty$ we remark that the right-hand side can be written in the following way:

$$\mathcal{C}(p, q; i, j; t) = \frac{\mu}{(2i\pi)^3} \int_C \lambda^t d\lambda \sum_{\alpha=1}^{4} \int_{C_1}\int_{C_2} \frac{dz_1 dz_2 \delta_i^\alpha \delta_\alpha^j}{z_1^{p+1} z_2^{q+1}(\lambda - \lambda_\alpha)}, \qquad \text{(C1)}$$

where C is a circle in the λ-complex plane centered at the origin and with a radius greater than one. The asymptotic part as $t \to \infty$ of the right-hand side of (C1) is given by the singularity in the λ-complex plane which is the closest to the unit circle. We note that the right-hand side of (C1) can be written as

$$\mathcal{C}(p, q; i, j; t) = \int_C \lambda^t d\lambda \int_{C_1}\int_{C_2} \frac{dz_1 dz_2}{z_1^{p+1} z_2^{q+1}} \frac{P(z_1, z_2; \lambda)}{S(z_1, z_2; \lambda)}, \qquad \text{(C2)}$$

where $P(z_1, z_2; \lambda)$ is some polynomial in $(z_1, z_2; \lambda)$ and $S(z_1, z_2; \lambda)$ is the characteristic polynomial of the matrix $M(z_1, z_2)$ introduced in Sec. 6:

$$S(z_1, z_2; \lambda) = \lambda^4 + (\mu - 1)(z_1 + z_2)(1 + \overline{z_1 z_2})\lambda^3 + (1 - 2\mu)[z_1 z_2 + (z_1 + z_2)^2 \overline{z_1 z_2} + \overline{z_1 z_2}]\lambda^2 + (3\mu - 1)(z_1 + z_2)(1 + \overline{z_1 z_2})\lambda + (1 - 4\mu).$$

It is known that a necessary condition for λ to be a singularity of the $z_1 z_2$ integral in (C2) is that

$$S(z_1, z_2, \lambda) = 0 \qquad \frac{\partial S}{\partial z_1}(z_1, z_2, \lambda) = 0, \qquad \frac{\partial S}{\partial z_2}(z_1, z_2, \lambda) = 0. \qquad \text{(C3)}$$

An elementary calculation shows that the only solutions of (C3) are $\lambda = \pm 1$, $\lambda = \pm\sqrt{1 - 4\mu}$, $\lambda = \pm(1 - 4\mu)$. From the above considerations it follows that it suffices to find the leading part of the singularity at $\lambda = \pm 1$ and in fact at $\lambda = +1$ (the contribution of $\lambda = -1$ is the same by a symmetry argument).

At this step it is more convenient to return to the expression (C1) and to use the possibility of restricting the integration domain as suggested by the formula (6.8). Let us investigate the contribution of each region.

(1) First we compute the contribution of the region $\mathcal{R}_1$ (see Fig. 4). The region $\mathcal{R}_1$ is a small region of $C_1 \times C_2$ surrounding the points $(z_1, z_2) = (1, 1)$ and $(-1, -1)$. A symmetry argument allows us to compute only the contribution of the small region around $(z_1, z_2) = (1, 1)$. We introduce the new arguments (k_1, k_2) such that $z_1 = \exp(i \arctan k_1)$, $z_2 = \exp(i \arctan k_2)$. A straightforward calculation leads to the Table CI for the expansion of the eigenvalues and eigenvectors of

$$M(z_1, z_2) \qquad \text{near } k = \sqrt{k_1^2 + k_2^2} = 0.$$

In this table we use the convention that $k_3 = -k_1, k_4 = -k_2$ and $o(k^i)$ means a quantity of order greater than k^i.

It is clear that the purely damped mode will give an exponentially decreasing contribution. Therefore, we have to consider only the contributions of the three modes $\lambda_D, \lambda_\pm$. To simplify the expressions, we introduce the following notation:

$$\Phi_{i,j}^{\alpha}(k_1, k_2) = \delta_i^\alpha \delta_\alpha^j (k_1, k_2)/(1 + k_1^2)(1 + k_2^2) z_1^p z_2^q.$$

Now the contribution to $(\mathcal{R}_1)$ of the mode λ_D reads as

$$(\mathcal{R}_1)_D \simeq \frac{\mu}{2\pi^2(2i\pi)} \int_C \lambda^t d\lambda \int_{k \leq \eta} dk_1 dk_2 \frac{\Phi_{i,j}^D(k_1, k_2)}{\lambda - 1 + \dfrac{(1 - 2\mu)}{2\mu}\dfrac{k_1^2 k_2^2}{k^2}} \qquad \text{(C4)}$$

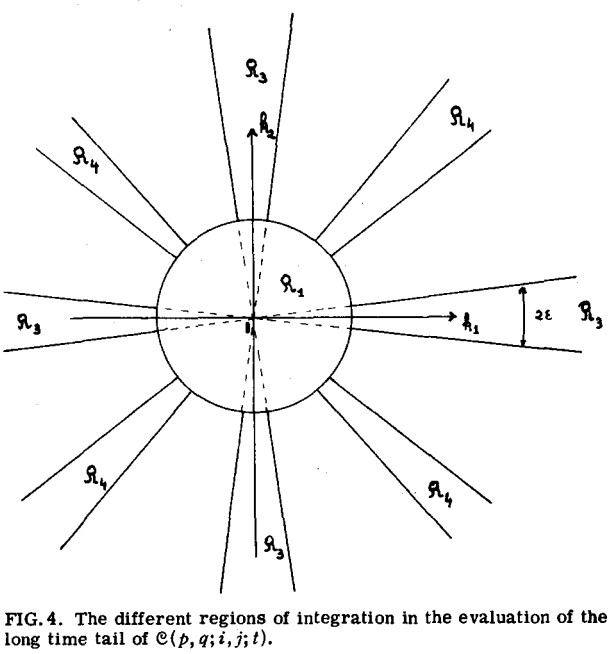


FIG. 4. The different regions of integration in the evaluation of the long time tail of $\mathcal{C}(p, q; i, j; t)$.

TABLE CI

Mode	Eigenvalue	Eigenvector
Vorticity diffusion	$\lambda_D = 1 - \frac{(1-2\mu)k_1^2k_2^2}{2\mu k^2} + o(k^2)$	$\delta_i^D = \frac{k_{i+1}}{\sqrt{2}k} + o(k)$
Sound waves	$\lambda_\pm = 1 \pm \frac{ik}{\sqrt{2}} + \frac{2(1-2\mu)k_1^2k_2^2 - (\mu+1/2)k^4}{8\mu k^2} + o(k^2)$	$\delta_i^\pm = \frac{1}{2}\left(\pm\frac{1}{\sqrt{2}} - \frac{k_i}{k}\right) + o(k)$
Purely damped mode	$\lambda_x = (1-4\mu) + o(k)$	$\delta_i^x = \frac{1}{2}(-1)^i + o(k)$

TABLE CII.

	$r=0$	$r=1$	$r>1$
$\int_0^\eta k^n dk \int_{-\epsilon}^{+\epsilon} \frac{\theta^{2r}d\theta}{\lambda - 1 + \frac{1}{4}\alpha(\sin 2\theta)^2k^2} \simeq$	$\frac{\pi}{\sqrt{\alpha}}\int_0^\eta k^{n-1}dk(\lambda-1)^{-1/2}$	$-\frac{\pi}{\alpha^{3/2}}\int_0^\eta k^{n-1}dk(\lambda-1)^{1/2} + cte(\lambda-1)^{(n-1)/2} \times \log(\lambda-1)$	$o(\lambda-1)^{1/2} + cte(\lambda-1)^{(n-1)/2} \times \log(\lambda-1)$

	n odd	n even
$\int_0^\eta k^n dk \int_\epsilon^{\pi/2-\epsilon} \frac{\theta^{2r}d\theta}{\lambda - 1 + \frac{1}{4}\alpha(\sin 2\theta)^2k^2} \simeq$	$cte \times (\lambda-1)^{(n-1)/2}\log(\lambda-1)$	analytic

TABLE CIII.

	$r=0$	$r=1$	$r>1$
$\int_{k\geq\eta} dk \int_{-\epsilon}^{+\epsilon} \frac{k^n\theta^{2r}d\theta}{\lambda - 1 - \alpha k^2\theta^2} \simeq$	$\frac{\pi}{\sqrt{\alpha}}\int_{k\geq\eta} k^{n-1}dk(\lambda-1)^{-1/2}$	$-\frac{\pi}{\alpha^{3/2}}\int_{k\geq\eta} k^{n-1}dk(\lambda-1)^{1/2}$	$o(\lambda-1)^{1/2}$

We see that the $k_1 \times k_2$ integral is singular in $\lambda = 1$. In order to find the leading part of the singularity, we shall separate the disk into sectors as indicated in Fig. 4. Let us set $k_1 = k\cos\theta$, $k_2 = k\sin\theta$. We can show that the contribution of the sector $\epsilon \leq \theta \leq \pi/2 - \epsilon$ is always negligeable in comparison with that of the sector $|\theta| \leq \epsilon$. Here we assume that $\epsilon \leq 1$. In fact we have Table CII giving the leading parts of the singularity at $\lambda = 1$ of typical integrals with which we are concerned. Similar results are obtained for the other sectors. Now, if we expand the integrand in the $k_1 \times k_2$ integral near $k = 0$ and $\theta = 0$, we find terms like $k^n\theta^{2r}$ with $r \leq n$. In view of Table CII it is clear that if $r < n$ the sectors $|\theta - q\pi/2| \leq \epsilon$ with $q \in \{1, 2, 3, 4\}$ give always the leading part of the singularity in $\lambda = 1$. For $r = n = 1$ an explicit calculation shows that the logarithmic parts of the contributions of the two sectors $|\theta| \leq \epsilon$ and $\epsilon \leq \theta \leq \pi/2 - \epsilon$ cancel, so that we have finally

$$(\mathcal{R}_1)_D \simeq \begin{cases} \frac{\mu}{2\pi}\left(\frac{2\mu}{1-2\mu}\right)^{1/2}\int_{-\eta}^{+\eta}\Phi^D_{i,j}(k_{1=0},k_2)dk_2 \times \frac{1}{2i\pi}\int_C \lambda^t d\lambda(\lambda-1)^{-1/2} & \text{if } i+j \text{ even}, \\ & i = 1, \\ \frac{\mu}{2\pi}\left(\frac{2\mu}{1-2\mu}\right)^{1/2}\int_{-\eta}^{+\eta}\Phi^D_{i,j}(k_1,k_{2=0})dk_1 \times \frac{1}{2i\pi}\int_C \lambda^t d\lambda(\lambda-1)^{-1/2} & \text{if } i+j \text{ even}, \\ & i = 2, \\ -\frac{\mu}{4\pi}\left(\frac{2\mu}{1-2\mu}\right)^{3/2}\left[\int_{-\eta}^{+\eta}\frac{\partial^2}{\partial\theta^2}\Phi^D_{i,j}|_{k_2=0}dk_1 + \int_{-\eta}^{+\eta}\frac{\partial^2}{\partial\theta^2}\Phi^D_{i,j}|_{k_1=0}dk_2\right] \times \frac{1}{2i\pi}\int_C \lambda^t d\lambda(\lambda-1)^{1/2} & \text{if } i+j \text{ odd}. \end{cases} \tag{C5}$$

The contribution to $(\mathcal{R}_1)$ of the sound waves are much simpler; in fact we have

$$(\mathcal{R}_1)_\pm \simeq \frac{\mu}{2\pi^2(2i\pi)}\int_C \lambda^t d\lambda \int_{k\leq\eta} dk_1dk_2 \frac{\Phi^\pm_{i,j}(k_1,k_2)}{\lambda - 1 \mp ik/\sqrt{2}}. \tag{C6}$$

The $k_1 \times k_2$ integral is again singular in $\lambda = 1$, but the leading part of the singularity is always weaker than $(\lambda-1)^{1/2}$. To see this, we remark that

$$\int_{k\leq\eta} dk_1dk_2 \frac{1}{\lambda - 1 \mp ik/\sqrt{2}} \simeq cte(\lambda-1)\log(\lambda-1) \quad \text{as } \lambda \simeq 1.$$

This shows that the contribution of the sound waves is always negligeable in comparison with the contribution of the mode λ_D.

(2) We compute now the contribution of the region $\mathcal{R}_2 \cup \mathcal{R}_3$. Symmetry considerations enable us to restrict our study in the neighborhood of $(z_1 = 1, z_2 \neq 1)$ or $(z_2 = 1, z_1 \neq 1)$. In this region only one mode has to be considered. This mode is the natural continuation of the mode λ_D. We use again the variables k_1, k_2 and take as the region of integration the sector $k \geq \eta$, $|\theta| \leq \epsilon$ and its image by rotation of angle $n\pi/2$, with $n \in \{1, 2, 3, 4\}$. In the following we shall calculate the contribution of the first sector $k \geq \eta$, $|\theta| \leq \epsilon$. The contributions of the others are obtained by similar arguments; we have

$$(\mathcal{R}_2) + (\mathcal{R}_3) \simeq \frac{\mu}{2\pi^2(2i\pi)}\int_C \lambda^t d\lambda \int_{k\geq\eta}\int_{|\theta|\leq\epsilon} dk_1dk_2 \frac{\Phi^D_{i,j}(k_1,k_2)}{\lambda - 1 + [(1-2\mu)/2\mu]k^2\theta^2}. \tag{C7}$$

After expanding the integrand of the $k_1 \times k_2$ integral around $\theta = 0$, we are led to evaluate the leading part of the singularity at $\lambda = 1$ of integrals of the type shown in Table CIII. Therefore, taking into account all the contributions of the same order for the four sectors, we get the following results:

$$(\mathcal{R}_2)+(\mathcal{R}_3)\simeq\begin{cases}\dfrac{\mu}{2\pi}\left(\dfrac{2\mu}{1-2\mu}\right)^{1/2}\int_{|k_2|\geq\eta}\Phi^D_{i,j}(k_{1=0},k_2)dk_2\times\dfrac{1}{2i\pi}\int_C d\lambda\,\lambda^t(\lambda-1)^{-1/2} & \text{if } i+j \text{ even},\ i=1,\\ \dfrac{\mu}{2\pi}\left(\dfrac{2\mu}{1-2\mu}\right)^{1/2}\int_{|k_1|\geq\eta}\Phi^D_{i,j}(k_1,k_{2=0})dk_1\times\dfrac{1}{2i\pi}\int_C d\lambda\,\lambda^t(\lambda-1)^{-1/2} & \text{if } i+j \text{ even},\ i=2,\\ -\dfrac{\mu}{4\pi}\left(\dfrac{2\mu}{1-2\mu}\right)^{3/2}\left[\int_{|k_1|\geq\eta}\dfrac{\partial^2}{\partial\theta^2}\Phi^D_{i,j}|_{k_2=0}dk_1+\int_{|k_2|\geq\eta}\dfrac{\partial^2}{\partial\theta^2}\Phi^D_{i,j}|_{k_1=0}dk_2\right]\times\dfrac{1}{2i\pi}\int_C d\lambda\,\lambda^t(\lambda-1)^{1/2} & \text{if } i+j \text{ is odd}\end{cases}\tag{C8}$$

(3) It remains to compute the contribution of $\mathcal{R}_4$. In this region we have to consider two modes which are the natural continuations of the sound waves. The contribution of these modes can be written as

$$(\mathcal{R}_4)=\frac{\mu}{(2i\pi)^3}\int_C\lambda^t d\lambda\left[\int_{\substack{k\geq\eta\\ z_1\neq z_2\\ z_1\neq\bar z_2}}\left(\frac{\Phi^+_{i,j}(z_1,z_2)}{\lambda-z_1}+\frac{\Phi^-_{i,j}(z_1,z_2)}{\lambda-\bar z_1}\right)dz_1dz_2\right].$$

It is seen that the $z_1\times z_2$ integral is not singular in $\lambda=1$.

Therefore, the long time tail of the correlation functions we have considered is given by adding the results (C. 5) and (C. 8) together. After this step a new feature appears. In fact we have for $i+j$ even and $i=1$, for instance,

$$\mathcal{C}(p,q;i,j;t)\simeq\frac{\mu}{4i\pi}\left(\frac{2\mu}{1-2\mu}\right)^{1/2}\int_{C_2}\frac{\delta^p_i\delta^j_b(z_1=1,z_2)}{z_2^{q+1}}dz_2\times\frac{1}{2i\pi}\int_C d\lambda\,\lambda^t(\lambda-1)^{-1/2}.\tag{C9}$$

But $\delta^p_i\delta^j_b(z_{1=1},z_2)$ is independent of z_2 (see Table CI) so that the right-hand side of (C9) is different from zero if and only if $q=0$. This means that if $q\neq 0$, one has to take the next term in the expansion of the integrand of the $k_1\times k_2$ integral. The final results are summarized in Table CIV.

APPENDIX D

We want to calculate the expectation value $\mathcal{C}(p,q,i;j;t)$ when $|p|+|q|=t$ and to compute the asymptotic part as $t\to\infty$ in the two following cases:

(a) p (or q) is fixed;

(b) $|p|-|q|$ is fixed.

We shall suppose that p and q are positive; the other cases lead to similar calculations. Let us treat the first case:

(a) p is fixed: We have the following equations:

$$\mathcal{C}(p;i,3;t+1)=(1-\mu)\mathcal{C}(p-1;i,3;t)+\mu\mathcal{C}(p-1;i,4;t)$$
$$\mathcal{C}(p;i,4;t+1)=\mu\mathcal{C}(p;i,3;t)+(1-\mu)\mathcal{C}(p;i,4;t),\tag{D1}$$

where we have set $\mathcal{C}(p;i,j;t)=\mathcal{C}(p,q;i,j;t)$ with $q=t-p$. Let us introduce the generating functions $\mathcal{C}(z_1;i,j;\lambda)$ such that:

$$\mathcal{C}(p;i,j;t)=\frac{1}{(2i\pi)^2}\int_C\lambda^{t-1}d\lambda\int_{C_1}\frac{dz_1}{z_1^{p+1}}\mathcal{C}(z_1;i,j;\lambda),$$

where C (resp. C_1) is the unit circle in the λ (resp. z_1) complex plane. Standard calculations lead to

$$\mathcal{C}(p;i,j;t)=\frac{1}{(2i\pi)^2}\int_C\lambda^{t-1}d\lambda\int_{C_1}\frac{dz_1}{z_1^{p+1}}\frac{P(z_1,\lambda;i,j)}{S(z_1,\lambda)},$$

where P is a polynomial of first degree with respect to λ and z_1 and $S(z_1,\lambda)$ is the determinant of the matrix

$$\begin{vmatrix}\lambda-(1-\mu)z_1 & -\mu z_1\\ -\mu & \lambda-(1-\mu)\end{vmatrix}.$$

The integral is performed by a calculus of residues at $z_1=\lambda[\lambda-(1-\mu)]/(1-\mu)[\lambda-(1-\mu)]+\mu^2$. Then the λ integral is similarly performed at $\lambda=(1-\mu)$. This leads after simple calculations to

$$\mathcal{C}(p;i,j;t)\simeq cte\times t^p(1-\mu)^t.\tag{D2}$$

(b) $p-q$ is fixed: Here it is convenient to introduce the argument $p'=p-q$ and to rewrite the recurrence relation in the following way:

$$\mathcal{C}(p';i,3;t+1)=(1-\mu)\mathcal{C}(p'+1;i,3;t')+\mu\mathcal{C}(p'+1;i,4;t),$$
$$\mathcal{C}(p';i,4;t+1)=\mu\mathcal{C}(p'-1;i,3;t)+(1-\mu)\mathcal{C}(p'-1,i,4;t).$$

Now, using the same technique as in the previous case, we have

$$\mathcal{C}(p';i,j;t)=\frac{1}{(2i\pi)^2}\int_C\lambda^{t-1}d\lambda\int_{C_1}\frac{dz_1}{z_1^{p'+1}}\frac{P(z_1,\lambda;i,j)}{S(z_1,\lambda)},\tag{D3}$$

where

$$S(z_1,\lambda)=\lambda(1-\mu)z_1^2-(\lambda^2+1-2\mu)z_1+\lambda(1-\mu).\tag{D4}$$

TABLE CIV.

$$\mathcal{C}(p,q;i,j;t)\simeq\begin{cases}\dfrac{\mu}{\pi}\left(\dfrac{2\mu}{1-2\mu}\right)^{1/2}\dfrac{\sqrt{\pi}}{2}\delta^p_i\delta^j_b(z_1=1)t^{-1/2} & \text{if } i+1 \text{ even}, i=1, q=0\\ \dfrac{\mu}{\pi}\left(\dfrac{2\mu}{1-2\mu}\right)^{1/2}\dfrac{\sqrt{\pi}}{2}\delta^p_i\delta^j_b(z_2=1)t^{-1/2} & \text{if } i+j \text{ even}, i=2, p=0\\ -\dfrac{\mu}{2\pi^2}\left(\dfrac{2\mu}{1-2\mu}\right)^{3/2}\dfrac{\sqrt{\pi}}{2}\left[\int\dfrac{\partial^2}{\partial\theta^2}\Phi^D_{i,j}|_{k_2=0}dk_1+\int\dfrac{\partial^2}{\partial\theta^2}\Phi^D_{i,j}|_{k_1=0}dk_2\right]t^{-3/2} & \text{in the other cases}\end{cases}$$

The z_1 integral in (D3) is performed by the calculus of residues at the root of (D4) which has a modulus smaller than one. We get a function of λ which is singular at $\lambda = \pm 1$ and $\lambda = \pm (1 - 2\mu)$. The asymptotic part of $\mathcal{C}(p', t; i, j)$ as $t \to \infty$ is thus given by the singularity at $\lambda = \pm 1$. An elementary calculation shows that if $i = 1$ or 2, the leading part of the singularity is $C \times \sqrt{\lambda \pm 1}$, and if $i = 3$ or 4, the leading part of the singularity is $C \times (\lambda \pm 1)^{-1/2}$. This gives immediately the estimates:

$$\mathcal{C}(p'; i, j; t) \simeq \begin{cases} C \times t^{-1/2} & \text{if } i = 3 \text{ or } 4, \\ C \times t^{-3/2} & \text{if } i = 1 \text{ or } 2. \end{cases}$$

[1]Ya. G. Sinai, Funct. Anal. Appl. **6**, 1, 41 (1972).
[2]O. de Pazzis, Commun. Math. Phys. **22**, 121 (1971).
[3]J. L. Lebowitz and J. K. Percus, Phys. Rev. **155**, 122 (1967)
[4]O. E. Lanford, Commun. Math. Phys. **11**, 257 (1969).
[5]O. de Pazzis, Commun. Math. Phys. (in press).
[6]J. Hardy and Y. Pomeau, J. Math. Phys. **13**, 7 (1972); J. Math. Phys. **13**, 1042 (1972).
[7]Y. Pomeau, Phys. Rev. A **3**, 1174 (1971).
[8]The authors are gratefully indebted to E. H. Hauge for suggesting to them this modification.
[9]D. Ruelle, *Statistical Mechanics* (Benjamin, New York, 1969).
[10]J. P. Conze, "Entropie d'un groupe abélien de transformations," preprint (1972).
[11]A. Arnold and A. Ayez, *Problèmes ergodiques de la mécanique classique* (Gauthiers-Villars, Paris, 1967).
[12]P. Halmos, *Measure Theory* (Van Nostrand, New York, 1950).
[13]M. Smorodinsky, *Erogodic Theory*, Entropy Lecture Notes in Mathematics (Springer-Verlag, Berlin, 1971).
[14]M. H. Ernst, E. H. Hauge, and Van Leeuwen, Phys. Rev. A **4**, 2055 (1971).

Duffin–Kemmer–Petiau subalgebras: Representations and applications*

Ephraim Fischbach

Physics Department, Purdue University, Lafayette, Indiana 47907

Michael Martin Nieto

Los Alamos Scientific Laboratory, University of California, Los Alamos, New Mexico 87544

C. K. Scott†

Physics Department, McMaster University, Hamilton, Ontario, Canada

(Received 22 March 1973)

A reduction of the Duffin–Kemmer–Petiau algebra to a direct sum of irreducible subalgebras for spin-0 and spin-1 bosons is presented. The subalgebras are defined by multiplication rules for the linearly independent basis elements. In the representations discussed the spin projection operators are independent basis elements of the subalgebras. The formal utility of these representations is demonstrated by obtaining the reduction of arbitrary operator products and trace theorems. The practical utility is demonstrated by application to the analysis of free and interacting boson field currents. Most importantly, one can understand the differences between DKP nonconserved currents and those obtained from second-order wave equations.

1. INTRODUCTION

In their original studies of first-order covariant wave equations describing spin-0 and spin-1 fields, Duffin,[1] Kemmer,[2] and Petiau[3] introduced a set of matrices that generate an algebra referred to as the Duffin–Kemmer–Petiau (DKP) algebra. Kemmer's development of the theory of the meson wave equation emphasized the similarity to the Dirac theory of the electron although the similarity is essentially only in the form of the equations. The algebra generated by the DKP matrices has a structure very different from that of the Dirac matrices. A notable difference is that the DKP equation simultaneously describes both spin-0 and spin-1 particles. That is, the algebra is reducible.

Many authors have discussed the structure of the DKP algebra for mesons. The original analyses of Fujiwara[4] and Tokuoka and Tanake[5] are the basis of the present discussion. These original studies introduced spin-projection operators which project from the DKP field the spin-0 and spin-1 components. More recently, Tokuoka[6] has used the spin projection operators to find the basis elements of the irreducible representations of the DKP meson algebra. In addition, he used the technique to find subsidiary algebraic relations which when combined with the equation defining the DKP algebra yield matrices which make up a particular irreducible representation. This approach has been formalized by Shimpuku[7] in a general analysis of DKP algebras using ring theory.

The analysis presented here can be regarded as an extension of the work of Tokuoka. We have found that past analyses of the reduction of the algebra have tended to emphasize the mathematical structure of the algebra rather than development of a formalism that is practical for the analysis of physical problems. It is in this latter direction that we extend the work of previous authors.

We first note that it is not always advantageous to work with the irreducible representations of the DKP meson algebra. For example, in certain formal considerations the reducible form of the algebra is most convenient since one can treat both spin-0 and spin-1 particles simultaneously. However, in consideration of particular situations involving either spin-0 or spin-1 particles the redundancy of the reducible form of the algebra obscures the actual structure of the problem. In addition, it is often of interest to compare results of an application of the DKP equation with results obtained from the Klein–Gordon (KG) or Proca equations, and such a comparison is not always obvious with the standard representations of the DKP algebra. As observed in the above paragraph, it is this latter direction in which we are most interested.

Since the spin projection operators introduced by Fujiwara[4] select from the sixteen component DKP field those components corresponding to spin-0 and spin-1, the projection operators provide a natural starting point for a physics-oriented reduction of the algebra. Moreover, the spin projection operators provide the natural formalism for demonstrating the equivalence of the *free field* DKP equations and the Klein–Gordon and Proca equations.[8] Thus, in line with the preceding arguments, *the essential point of our approach is to obtain a representation of the irreducible subalgebras in which the projection operators occur as independent basis elements.*

To facilitate the discussion of the subalgebras the necessary aspects of the DKP equation and the matrix algebra are reviewed in Sec. 2. Then a reduction of the DKP meson algebra to a direct sum of the irreducible subalgebras is presented in Sec. 3. In Secs. 4 and 5 we discuss useful properties of the DKP algebra and equation, such as the reduction of operator products, trace theorems, projection, raising and lowering operators and the consequent equations. We conclude our analysis in Secs. 6 and 7 by demonstrating the utility of these particular representations in applications to physical problems, involving spin-0 and spin-1 mesons, respectively.

2. THE DUFFIN-KEMMER-PETIAU EQUATION AND DKP ALGEBRA

The DKP equation for spin-0 and spin-1 mesons is

$$\partial_\lambda \beta_\lambda \psi(\mathbf{x}, t) = -m\psi(\mathbf{x}, t). \tag{1}$$

Important Note: For wave and current equations such as (1) above, we will use the summation convention for repeated Lorentz indices. However, when dealing with the DKP algebra elements above, summation will only be implied when a $\sum$ is explicitly written. See, for

example, the difference between equations (7) and (23) below.

The four matrices β_λ which are defined by the algebraic relation

$$\beta_\mu\beta_\nu\beta_\lambda + \beta_\lambda\beta_\nu\beta_\mu = \beta_\mu\delta_{\nu\lambda} + \beta_\lambda\delta_{\nu\mu} \tag{2}$$

generate the DKP meson algebra. In detail the defining algebraic relations are

$$\beta_\mu\beta_\nu\beta_\mu = \beta_\mu\delta_{\nu\mu}, \tag{3a}$$

$$\beta_\mu\beta_\nu^2 = (1-\beta_\nu^2)\beta_\mu, \qquad \mu \neq \nu, \tag{3b}$$

$$\beta_\mu\beta^2 = (5-\beta^2)\beta_\mu, \qquad \beta^2 = \sum_\lambda \beta_\lambda\beta_\lambda. \tag{3c}$$

From Eq. (3b) it follows that

$$\beta_\mu^2\beta_\nu^2 = \beta_\nu^2\beta_\mu^2. \tag{4}$$

Using Eq. (2) it is a straightforward task to write down the 126 independent elements of the DKP algebra. To do so it is convenient to introduce auxiliary elements associated with each index. There are two convenient ways of introducing the auxiliary elements. The first way is to define

$$\eta_\mu = 2\beta_\mu^2 - 1, \tag{5a}$$

$$\eta_5 = \eta_1\eta_2\eta_3\eta_4, \tag{5b}$$

and associate with each value of the index λ the triplet $(1, \beta_\lambda, \eta_\lambda)$. The second way is to define

$$\xi_\mu^+ = \beta_\mu^2, \qquad \xi_\mu^- = 1-\beta_\mu^2, \tag{6}$$

and with each value of the index λ associate the triplet $(\beta_\lambda, \xi_\lambda^+, \xi_\lambda^-)$. The algebra of the triplets $(1, \beta, \eta)$ is given by Eq. (2) and the following easily verified relations:

$$\eta_\lambda^2 = 1, \tag{7a}$$

$$\eta_\lambda\eta_\mu = \eta_\mu\eta_\lambda, \tag{7b}$$

$$\eta_5\eta_\lambda = \eta_\lambda\eta_5, \tag{7c}$$

$$\eta_\lambda\beta_\mu = -\beta_\mu\eta_\lambda \quad (\mu \neq \lambda), \tag{7d}$$

$$\beta_\mu\eta_\mu = \eta_\mu\beta_\mu = \beta_\mu, \tag{7e}$$

$$\eta_5\beta_\lambda = -\beta_\lambda\eta_5. \tag{7f}$$

Likewise the algebra of the triplets (β, ξ^+, ξ^-) is given by Eq. (2) and the following relations:

$$\xi_\lambda^+ + \xi_\lambda^- = 1, \tag{8a}$$

$$(\xi_\lambda^\pm)^2 = \xi_\lambda^\pm, \tag{8b}$$

$$\xi_\lambda^+\xi_\lambda^- = \xi_\lambda^-\xi_\lambda^+ = 0, \tag{8c}$$

$$\xi_\lambda^+\xi_\nu^- = \xi_\nu^-\xi_\lambda^+, \tag{8d}$$

$$\xi_\lambda^+\xi_\nu^+ = \xi_\nu^+\xi_\lambda^+, \tag{8e}$$

$$\xi_\lambda^-\xi_\nu^- = \xi_\nu^-\xi_\lambda^-, \tag{8f}$$

$$\beta_\lambda\xi_\lambda^+ = \xi_\lambda^+\beta_\lambda = \beta_\lambda, \tag{8g}$$

$$\beta_\lambda\xi_\lambda^- = \xi_\lambda^-\beta_\lambda = 0, \tag{8h}$$

$$\beta_\lambda\xi_\nu^+ = \xi_\nu^-\beta_\lambda, \qquad \nu \neq \lambda, \tag{8i}$$

$$\beta_\lambda\xi_\nu^- = \xi_\nu^+\beta_\lambda, \qquad \nu \neq \lambda. \tag{8j}$$

It follows from the definitions (5) and (6) that the two sets of triplets are related by

$$\xi_\lambda^+ = \tfrac{1}{2}(1+\eta_\lambda), \qquad \xi_\lambda^- = \tfrac{1}{2}(1-\eta_\lambda), \tag{9}$$

and one has the additional algebraic relations

$$\eta_\lambda\xi_\lambda^\pm = \pm\,\xi_\lambda^\pm, \tag{10a}$$

$$[\eta_5, \xi_\lambda^\pm] = 0. \tag{10b}$$

The 126 linearly independent elements of the algebra are listed in Table I. The independent elements listed in Table I are for the algebra generated by the four matrices β_μ and the unity (I). One can also consider the subalgebra generated by the β matrices alone.[9] The subalgebra has 125 linearly independent elements.

As is well known, the DKP algebra is reducible to irreducible subalgebras of 1, 25, and 100 linearly independent elements corresponding to the trivial identically vanishing, spin-0, and spin-1 equations, respectively. To obtain the linearly independent elements of the subalgebras Tokuoka[6] used a theorem of Harish-Chandra[10] to write down the unit element for each of the subalgebras. With the unit elements known one can then project out from the full algebra the linearly independent basis elements of the subalgebras.

In the spirit of our above discussion we will first obtain the subalgebras in terms of the spin projection operators defined by Fujiwara.[4] Defining

$$P = \beta_1^2\beta_2^2\beta_3^2\beta_4^2, \qquad P_\nu = P\beta_\nu, \qquad {}_\nu P \equiv (P_\nu)^\dagger = \beta_\nu(P) \tag{11}$$

it follows from the algebra of the β matrices that

$$P_\nu\beta_\lambda = P\delta_{\nu\lambda}. \tag{12}$$

From (11) and (12) and the DKP equation (1) we have that

$$\partial_\mu P\psi = -mP_\mu\psi \quad \text{and} \quad \partial_\mu P_\mu\psi = -mP\psi. \tag{13}$$

TABLE I: The 126 independent elements of the DKP meson algebra. The elements Γ_λ can be taken to be one of either η_λ, ξ_λ^+, or ξ_λ^-. For example, Kemmer[2] used the case $\Gamma_\lambda = \eta_\lambda$.

Element	Number of Elements
I	1
β_μ	4
$\beta_\mu\beta_\nu$	12
$\beta_\mu\beta_\nu\beta_\rho$	12
$\beta_\mu\beta_\nu\beta_\rho\beta_\sigma$	6
Γ_μ	4
$\Gamma_\mu\Gamma_\nu$	6
$\Gamma_\mu\Gamma_\nu\Gamma_\rho$	4
$\Gamma_\mu\Gamma_\nu\Gamma_\rho\Gamma_\sigma$	1
$\Gamma_\mu\beta_\nu$	12
$\Gamma_\mu\beta_\nu\beta_\rho$	24
$\Gamma_\mu\beta_\nu\beta_\rho\beta_\sigma$	12
$\Gamma_\mu\Gamma_\nu\beta_\rho$	12
$\Gamma_\mu\Gamma_\nu\beta_\rho\beta_\sigma$	12
$\Gamma_\mu\Gamma_\nu\Gamma_\rho\beta_\sigma$	4
	126

The quantities $P\psi$ and $P_\mu\psi$ transform as spin-0 and spin-1 fields, respectively. Thus, the five-component DKP field describing spin-0 particles consists of a scalar field and a 4-vector field which is the 4-gradient of the scalar field. To obtain the spin-1 DKP field we define

$$R_\mu = \begin{cases} -\beta_1^2\beta_2^2\beta_3^2\beta_\mu\beta_4, & \mu = 1, 2, 3, \\ \beta_1^2\beta_2^2\beta_3^2(1-\beta_4^2), & \mu = 4, \end{cases} \tag{14}$$

$$R_{\mu\nu} = R_\mu\beta_\nu.$$

From the algebra of the β matrices it follows that

$$\begin{aligned} &R_{\mu\nu} = -R_{\nu\mu}, \\ &R_{\mu\nu}\beta_\lambda = R_\mu\delta_{\nu\lambda} - R_\nu\delta_{\mu\lambda}. \end{aligned} \tag{15}$$

Multiplying R_ν on to the left side of the DKP equation, we have

$$\begin{aligned} &\partial_\mu R_{\nu\mu}\psi = -mR_\nu\psi, \\ &\partial_\lambda R_\nu\psi - \partial_\nu R_\lambda\psi = -mR_{\nu\lambda}\psi. \end{aligned} \tag{16}$$

By defining

$$\begin{aligned} &Q_\nu = R_\nu\psi m^{-1/2}, \\ &U_{\lambda\nu} = \partial_\lambda Q_\nu - \partial_\nu Q_\lambda, \end{aligned} \tag{17}$$

Eqs. (16) become the usual spin-1 field equations:

$$\begin{aligned} &U_{\mu\nu} = \partial_\mu Q_\nu - \partial_\nu Q_\mu, \\ &\partial_\mu U_{\mu\nu} = m^2 Q_\nu. \end{aligned} \tag{18}$$

From Eq. (18) we see that the ten-component Kemmer field for the spin-1 particle consists of four components of a vector field (Q_μ) plus six components which are the field strengths ($U_{\mu\nu}$). [Note that when $m \to 0$, Eqs. (18) become Maxwell's equations with $U_{\mu\nu} \to F_{\mu\nu}$ and $Q_\mu \to A_\mu$.]

In the next section we will obtain representations of the spin-0 and spin-1 subalgebras in which the spin projection operators are basis elements of the subalgebras.

3. THE DKP SUBALGEBRAS

From Eqs. (11)–(15) it is apparent that the projection operators (P, P_μ) and $(R_\mu, R_{\mu\nu})$ will generate two right ideals of the algebra. Moreover, since

$$P(R_\mu) = (R_\mu)P = 0, \tag{19a}$$

$$(P_\mu)(R_\nu) = (R_\mu)(P_\nu) = 0, \tag{19b}$$

$$P(R_{\mu\nu}) = (R_{\mu\nu})P = 0, \tag{19c}$$

$$P_\lambda(R_{\mu\nu}) = (R_{\mu\nu})(P_\lambda) = 0, \tag{19d}$$

the two ideals have zero intersection. By including the Hermitian conjugate projection operators, which project the various components from the sixteen-component Hermitian adjoint DKP field, one obtains sets of elements which generate ideals of the algebra. The elements of the two ideals are the subalgebras corresponding to spin-0 and spin-1. That is, the β-algebra has been reduced to a direct sum of two irreducible subalgebras (references to the trivial subalgebra will only be made where necessary). The article by Shimpuku[7] is an excellent source of material regarding the formal proof of this decomposition. Consequently, our discussion will only indicate the formal arguments that are applicable.

A. Spin-0 subalgebra

The twenty-five elements of the spin-0 subalgebra (P algebra) are $\{P, P_\mu, {}_\mu P, {}_\mu P_\nu\}$, where we have included the sixteen elements

$${}_\mu P_\nu \equiv ({}_\mu P)(P_\nu) \tag{20}$$

in addition to the projection operators. It is a straightforward exercise to write down the multiplication table for the P-algebra and it is tabulated in Appendix 1. However, by extending the range of the indices to five instead of four the multiplication rule for the elements can be written in a compact form. Defining

$${}_5P_5 \equiv P, \quad {}_5P_\mu \equiv P_\mu, \quad {}_\mu P_5 \equiv {}_\mu P, \tag{21}$$

we have that the P-algebra consists of the elements $\{{}_aP_b \mid a, b = 1, \ldots, 5\}$. The multiplication rule is then

$$({}_aP_b)({}_cP_d) = {}_aP_d\delta_{bc}. \tag{22}$$

From Eq. (22) it follows that the twenty-five elements are linearly independent and form a basis for the spin-0 subalgebra. Representations of the P-algebra are given in Appendix 2.

The unit element of the P-algebra is

$$e_P = P + \sum_\mu ({}_\mu P_\mu). \tag{23}$$

With the unit element one can project from the full β-algebra the twenty-five independent elements of the five-dimensional spin-0 subalgebra. For example, from Eqs. (23) and (12) we have that

$$e_P\beta_\lambda = \beta_\lambda e_P = P_\lambda + {}_\lambda P. \tag{24}$$

The auxiliary elements for each index defined in Sec. 2 are

$$\eta_\lambda = P + {}_\lambda P_\lambda - \sum_{\mu\neq\lambda} ({}_\mu P_\mu), \tag{25a}$$

$$\xi_\lambda^+ = P + {}_\lambda P_\lambda, \tag{25b}$$

$$\xi_\lambda^- = \sum_{\mu\neq\lambda} ({}_\mu P_\mu). \tag{25c}$$

By repeated application of this procedure one obtains a set of relations between the elements of the P-algebra and the elements of the β-algebra in the five-dimensional representation.

The relation between the P-algebra and the β-algebra in the five-dimensional representation is given by

$$P = \beta_\mu^2\beta_\nu^2, \quad \mu \neq \nu, \tag{26a}$$

$$P_\mu = \beta_\nu^2\beta_\mu, \quad \mu \neq \nu, \tag{26b}$$

$${}_\mu P = \beta_\mu\beta_\nu^2, \quad \mu \neq \nu, \tag{26c}$$

$${}_\mu P_\mu = \beta_\mu^2(1-\beta_\nu^2), \quad \mu \neq \nu, \tag{26d}$$

$${}_\mu P_\nu = \beta_\mu\beta_\nu, \quad \mu \neq \nu. \tag{26e}$$

Equation (26) reflects the redundancy of the independent elements of the β-algebra when only a particular irreducible representation is considered. In particular, Eqs.

(26a)–(26d) are valid for *all* ν such that $\nu \neq \mu$, and so the redundancy of using the complete algebra to describe the spin-0 subalgebra is clearly exhibited. In addition to Eqs. (24), (25) and (26) there are the useful relations

$$\beta_\nu^2\beta_\mu - \beta_\mu\beta_\nu^2 = \eta_\nu\beta_\mu = P_\mu - {}_\mu P \quad (\mu \neq \nu), \tag{27a}$$

$$\sum_\nu(\beta_\nu^2\beta_\mu - \beta_\mu\beta_\nu^2) = 3(P_\mu - {}_\mu P), \tag{27b}$$

$$\beta^2 = \sum_\nu(\beta_\nu^2) = 3P + e_P, \tag{27c}$$

$$\eta_5 = 2P - e_P. \tag{27d}$$

We then have for the spin-0 subalgebra that the set of elements $\{{}_aP_b \mid a, b = 1, \ldots, 5\}$ defined by the multiplication rule

$$({}_aP_b)({}_cP_d) = {}_aP_d\delta_{bc} \tag{22}$$

and the definition

$$\beta_\lambda \equiv P_\lambda + {}_\lambda P \quad (\lambda = 1, 2, 3, 4) \tag{28}$$

replace Eq. (2) for defining the β-matrices and their algebra.

B. Spin-1 subalgebra

To obtain the spin-1 subalgebra we proceed in an analogous manner. From the definition of the projection operators in Eq. (14) and their Hermitian conjugates the multiplicative properties of the operators are given by

$$(R_\mu)(R_\nu) = R_\nu\delta_{\mu 4}, \tag{29a}$$

$$(R_\mu)(R_{\nu\lambda}) = R_{\nu\lambda}\delta_{\mu 4}, \tag{29b}$$

$$(R_{\mu\lambda})(R_\nu) = (R_{\mu\lambda})(R_{\nu\sigma}) = 0, \tag{29c}$$

$$(R_\mu)({}_\nu R) = R_4\delta_{\mu\nu}, \tag{29d}$$

$$(R_{\mu\nu})({}_\lambda R) = (R_\mu)({}_{\nu\lambda}R) = 0, \tag{29e}$$

$$(R_{\mu\nu})({}_{\lambda\sigma}R) = R_4\Delta_{\mu\nu\lambda\sigma}, \tag{29f}$$

where

$$\Delta_{\mu\nu\lambda\sigma} \equiv \delta_{\mu\sigma}\delta_{\nu\lambda} - \delta_{\mu\lambda}\delta_{\nu\sigma}. \tag{30}$$

So that the elements of the subalgebra can be given in a compact form, we introduce the matrices

$${}_\mu V_\nu \equiv ({}_\mu R)(R_\nu). \tag{31}$$

This is the set of elements

$$\begin{aligned} &{}_4V_4 = R_4 = {}_4R, \\ &{}_4V_k = R_k, \quad {}_kV_4 = {}_kR, \quad k = 1, 2, 3, \\ &{}_kV_j = ({}_kR)(R_j), \quad k, j = 1, 2, 3. \end{aligned} \tag{32}$$

The remaining elements of the subalgebra are

$$\begin{aligned} &{}_\mu V_{\nu\lambda} = ({}_\mu V_\nu)\beta_\lambda = ({}_\mu R)(R_{\nu\lambda}), \\ &{}_{\nu\lambda}V_\mu = \beta_\nu({}_\lambda V_\mu) = ({}_{\nu\lambda}R)(R_\mu), \\ &{}_{\nu\lambda}V_{\mu\sigma} = \beta_\nu({}_\lambda V_\mu)\beta_\sigma = ({}_{\nu\lambda}R)(R_{\mu\sigma}). \end{aligned} \tag{33}$$

Note that the elements are antisymmetric in double left and/or right indices. That is,

$$\begin{aligned} &{}_\mu V_{\nu\lambda} = -{}_\mu V_{\lambda\nu}, \quad {}_{\nu\lambda}V_\mu = -{}_\lambda V_\mu \\ &{}_{\nu\lambda}V_{\mu\sigma} = -{}_\lambda V_{\mu\sigma} = {}_\lambda V_{\sigma\mu}. \end{aligned} \tag{34}$$

The products of the sets of elements defined in Eq. (33) with elements of the β-algebra are given by

$$\begin{aligned} &({}_\mu V_{\nu\lambda})\beta_\sigma = {}_\mu V_\nu\delta_{\lambda\sigma} - {}_\mu V_\lambda\delta_{\nu\sigma}, \\ &\beta_\sigma({}_{\nu\lambda}V_\mu) = {}_\lambda V_\mu\delta_{\sigma\nu} - {}_\nu V_\mu\delta_{\sigma\lambda}, \end{aligned} \tag{35}$$

and the corresponding expressions for the set of elements ${}_{\nu\lambda}V_{\mu\sigma}$.

The one hundred linearly independent elements of the ten-dimensional spin-1 subalgebra (V-algebra) are thus the set

$$\{{}_\mu V_\nu, {}_\mu V_{\nu\lambda}, {}_{\nu\lambda}V_\mu, {}_{\nu\lambda}V_{\mu\sigma}\}.$$

The multiplication table for this set of elements is given in Appendix 1. However, as before, by extending the range of the indices the multiplication table can be summarized in a compact form. The elements of the V-algebra is then the set

$$\{{}_{ab}V_{cd} \mid a, b, c, d = 1, \ldots, 5\},$$

where

$${}_{ab}V_{cd} = -{}_{ba}V_{cd} = {}_{ba}V_{dc} \tag{36}$$

and

$$\begin{aligned} &{}_{5\mu}V_{\nu 5} = {}_\mu V_\nu, \\ &{}_{5\mu}V_{\nu\lambda} = {}_\mu V_{\nu\lambda}, \quad {}_{\nu\lambda}V_{\mu 5} = {}_{\nu\lambda}V_\mu. \end{aligned} \tag{37}$$

The multiplication rule for the V-algebra elements is then

$$({}_{ab}V_{cd})({}_{ef}V_{gh}) = {}_{ab}V_{gh}\Delta_{cdef}. \tag{38}$$

From the multiplication rule it follows immediately that the elements we have written down are in fact linearly independent and form an appropriate basis for the spin-1 subalgebra. A representation of the subalgebra is given in Appendix 2.

It is worth noting at this point that, for any element from the P-algebra and any element from the V-algebra,

$$\{P\}\{V\} = \{V\}\{P\} = 0. \tag{39}$$

The unit element of the V-algebra is

$$e_V = \sum_\mu({}_\mu V_\mu) + \frac{1}{2}\sum_{\mu\nu}({}_{\mu\nu}V_{\nu\mu}). \tag{40}$$

As for the P-algebra, the unit element e_V may be used to project the independent elements of the spin-1 subalgebra from the entire β algebra. For example,

$$e_V\beta_\lambda = \beta_\nu e_V = \sum_\mu({}_\mu V_{\mu\lambda} + {}_{\lambda\mu}V_\mu). \tag{41}$$

With repeated application of this procedure, one obtains the following relations between the V-algebra elements and the elements of the β-algebra [in all the relations (42) the indices are such that $\sigma \neq \lambda \neq \rho \neq \tau$]:

$${}_\tau V_\tau = \beta_\sigma^2\beta_\lambda^2\beta_\rho^2, \tag{42a}$$

$${}_\tau V_\rho = -\beta_\sigma^2\beta_\lambda^2\beta_\rho\beta_\tau, \tag{42b}$$

$${}_\tau V_{\tau\rho} = \beta_\sigma^2\beta_\lambda^2\beta_\rho, \tag{42c}$$

$$_{\rho\tau}V_{\tau} = \beta_\rho\beta_\sigma^2\beta_\lambda^2, \tag{42d}$$

$$_{\sigma}V_{\rho\lambda} = \beta_\tau^2\beta_\lambda\beta_\sigma\beta_\rho, \tag{42e}$$

$$_{\rho\lambda}V_{\sigma} = \beta_\lambda\beta_\sigma\beta_\rho\beta_\tau^2, \tag{42f}$$

$$_{\rho\tau}V_{\tau\rho} = \beta_\rho\beta_\sigma^2\beta_\lambda^2\beta_\rho, \tag{42g}$$

$$_{\lambda\tau}V_{\tau\sigma} = (1-\beta_\rho^2)\beta_\lambda\beta_\sigma, \tag{42h}$$

$$_{\rho\lambda}V_{\sigma\tau} = \beta_\lambda\beta_\sigma\beta_\rho\beta_\tau. \tag{42i}$$

In addition, there are the following useful relations:

$$\beta_\lambda^2 = \sum_\mu (_\mu V_\mu + {}_{\lambda\mu}V_{\mu\lambda}) - {}_\lambda V_\lambda,$$
$$\sum_\lambda \beta_\lambda^2 = 2e_V + \sum_\mu (_\mu V_\mu). \tag{43}$$

To summarize, we have shown that the spin-1 subalgebra is given by the set of one hundred linearly independent basis elements $\{_{ab}V_{cd} \mid a, b, c, d = 1, 2, 3, 4, 5 \mid V \text{ antisymmetric in } ab \text{ and/or } cd\}$. The elements are defined by the multiplication rule

$$(_{ab}V_{cd})(_{ef}V_{gh}) = {}_{ab}V_{gh}\Delta_{cdef}. \tag{44}$$

With ten-dimensional β matrices defined by

$$\beta_\lambda = \sum_\mu (_\mu V_{\mu\lambda} + {}_{\lambda\mu}V_\mu), \tag{45}$$

Eqs. (44) and (45) replace Eq. (2) in defining the β matrices and their algebra for the case of spin 1.

4. REDUCTION OF OPERATOR PRODUCTS AND TRACE THEOREMS

A. Operator products

We begin this section by considering the reduction of operator products involving the β matrices. A product of β matrices can be reduced to linear combinations of the 126 independent basis elements of the full β-algebra by using the commutation relation (2) for the β matrices. Our goal is to obtain the reduction as linear combinations of the basis elements of the irreducible subalgebras for spin-0 and spin-1. To perform the reduction in a systematic fashion we first introduce an auxiliary notation for the β matrices as linear combinations of the basis elements of the subalgebras.

For spin-0 and spin-1 we write

$$\beta_\lambda = (\beta_\lambda)_R + (\beta_\lambda)_L, \tag{46}$$

with

$$(\beta_\lambda)_R \equiv P_\lambda, \quad (\beta_\lambda)_L \equiv {}_\lambda P \tag{47a}$$

for spin-0, and

$$(\beta_\lambda)_R \equiv \sum_\mu {}_\mu V_{\mu\lambda}, \quad (\beta_\lambda)_L \equiv \sum_\mu {}_{\lambda\mu}V_\mu \tag{47b}$$

for spin-1. From the multiplication tables for the P- and V-algebras we have for both spin-0 and spin-1

$$(\beta_\lambda)_R(\beta_\sigma)_R = (\beta_\lambda)_L(\beta_\sigma)_L = 0, \tag{48}$$

while for spin-0

$$(\beta_\lambda)_R(\beta_\sigma)_L = P\delta_{\lambda\sigma}, \tag{49a}$$

$$(\beta_\lambda)_L(\beta_\sigma)_R = {}_\lambda P_\sigma, \tag{49b}$$

and for spin-1

$$(\beta_\lambda)_R(\beta_\sigma)_L = \sum_{\mu\nu} {}_\mu V_\nu \Delta_{\mu\lambda\sigma\nu} \tag{50a}$$

$$(\beta_\lambda)_L(\beta_\sigma)_R = \sum_{\mu\nu} {}_{\lambda\mu}V_{\nu\sigma}\delta_{\mu\nu} = \sum_\mu (_{\lambda\mu}V_{\mu\sigma}). \tag{50b}$$

Now consider the product of β matrices

$$\pi_n = \prod_{i=1}^{n} \beta_{\lambda_i}. \tag{51}$$

From the fact that $(\beta_\lambda)_R$ and $(\beta_\lambda)_L$ are nilpotent, we have for both spin-0 and spin-1

$$\pi_{2n} = (\beta_{\lambda_1})_R(\beta_{\lambda_2})_L(\beta_{\lambda_3})_R \cdots (\beta_{\lambda_{2n}})_L + (\beta_{\lambda_1})_L(\beta_{\lambda_2})_R \cdots (\beta_{\lambda_{2n}})_R \tag{52a}$$

and

$$\pi_{2n+1} = \pi_{2n}(\beta_{\lambda_{2n+1}})_R + \pi_{2n}(\beta_{\lambda_{2n+1}})_L. \tag{52b}$$

With π_n in the form of Eqs. (52), Eqs. (49) and (50) can be used to obtain the desired reduction. The results are:

Spin-0

$$\pi_1 = P_{\lambda_1} + {}_{\lambda_1}P, \tag{53a}$$

$$\pi_{2n+1} = {}_{\lambda_1}P\left(\prod_{i=1}^{n}\delta(\lambda_{2i}, \lambda_{2i+1})\right) + P_{\lambda_{2n+1}}\left(\prod_{i=1}^{n}\delta(\lambda_{2i-1}, \lambda_{2i})\right), \tag{53b}$$

$$\pi_{2n} = P\left(\prod_{i=1}^{n}\delta(\lambda_{2i-1}, \lambda_{2i})\right) + {}_{\lambda_1}P_{\lambda_{2n}}\left(\prod_{i=1}^{n-1}\delta(\lambda_{2i}, \lambda_{2i+1})\right). \tag{53c}$$

Spin-1

$$\pi_1 = \sum_{\mu_1} (_{\mu_1}V_{\mu_1\lambda_1} + {}_{\lambda_1\mu_1}V_{\mu_1}), \tag{54a}$$

$$\pi_2 = \sum_{\mu_1,\mu_2} [_{\mu_1}V_{\mu_2}\Delta(\mu_1, \lambda_1, \lambda_2, \mu_2) + {}_{\lambda_1\mu_1}V_{\mu_2\lambda_2}\delta(\mu_1, \mu_2)], \tag{54b}$$

$$\pi_3 = \sum_{\mu_1,\mu_2,\mu_3} [_{\mu_1}V_{\mu_3\lambda_3}\Delta(\mu_1, \lambda_1, \lambda_2, \mu_2)\delta(\mu_2, \mu_3) + {}_{\lambda_1\mu_1}V_{\mu_3}\delta(\mu_1, \mu_2)\Delta(\mu_2, \lambda_2, \lambda_3, \mu_3)], \tag{54c}$$

$$\pi_{2n}(n \geq 2) = \sum_\mu (_{\mu_1}V_{\mu_{2n}})\Delta(\mu_1, \lambda_1, \mu_2, \lambda_2) \times\left(\prod_{i=2}^{n}\Delta(\mu_{2i-2}, \lambda_{2i-1}, \lambda_{2i}, \mu_{2i})\right) + \sum_\mu (_{\lambda_1\mu_1}V_{\mu_{2n}\lambda_{2n}})\delta(\mu_1, \mu_2) \times\left(\prod_{i=2}^{n}\Delta(\mu_{2i-2}, \lambda_{2i-2}, \lambda_{2i-1}, \mu_{2i})\right), \tag{54d}$$

$$\pi_{2n+1}(n \geq 2) = \sum_\mu (_{\mu_1}V_{\mu_{2n+1}\lambda_{2n+1}})\Delta(\mu_1, \lambda_1, \lambda_2, \mu_2)\delta(\mu_{2n}, \mu_{2n+1}) \times\left(\prod_{i=2}^{n}\Delta(\mu_{2i-2}, \lambda_{2i-1}, \lambda_{2i}, \mu_{2i})\right) + \sum_\mu (_{\lambda_1\mu_1}V_{\mu_{2n+1}})\delta(\mu_1, \mu_2)\Delta(\mu_{2n}, \lambda_{2n}, \lambda_{2n+1}, \mu_{2n+1}) \times\left(\prod_{i=2}^{n}\Delta(\mu_{2i-2}, \lambda_{2i-2}, \lambda_{2i-1}, \mu_{2i})\right). \tag{54e}$$

In Eqs. (54d) and (54e) the summation is over all subscripts μ_i that occur in the equations. With the general expressions given in Eqs. (53) and (54) any product of β matrices can be reduced to a linear combination of two of the basis elements from each of the respective subalgebras. Incidentally, by using Eqs. (26) and (42) in combination with Eqs. (53) and (54) the reduction of a product to a linear combination of independent basis elements of the full β-algebra can be obtained.

Other operator products involving β matrices that are of interest contain the matrices β_μ contracted with some 4-vector a_μ. To treat this case we adopt the Feynman slash notation

$$\not{a} \equiv \beta\cdot a = a\cdot\beta = \sum_\mu \beta_\mu a_\mu. \tag{55}$$

For consistency we also use the slash notation for the contraction of β matrices with themselves

$$\not{\beta} \equiv \beta\cdot\beta = \sum_\mu \beta_\mu\beta_\mu. \tag{56}$$

The commutation relations for the slashed operators follow from the basic commutation relation (2). The relations for operators $\not{a}$ are

$$\not{a}\beta_\nu\beta_\lambda + \beta_\lambda\beta_\nu\not{a} = \not{a}\delta_{\nu\lambda} + \beta_\lambda a_\nu, \tag{57a}$$

$$\not{a}\not{b}\beta_\lambda + \beta_\lambda\not{b}\not{a} = \not{a}b_\lambda + \beta_\lambda b\cdot a, \tag{57b}$$

$$\not{a}\not{b}\not{c} + \not{c}\not{b}\not{a} = \not{a}b\cdot c + \not{c}b\cdot a. \tag{57c}$$

For the operator $\not{\beta}$ the basic commutation relations are

$$\not{\beta}\beta_\nu + \beta_\nu\not{\beta} = 5\beta_\nu, \tag{58a}$$

$$\not{\beta}\not{a} + \not{a}\not{\beta} = 5\not{a}. \tag{58b}$$

Useful relations that follow from (58a) and (58b) are

$$\not{\beta}\beta_\nu\beta_\mu - \beta_\nu\beta_\mu\not{\beta} = 0, \tag{59a}$$

$$\not{\beta}\not{a}\not{b} - \not{a}\not{b}\not{\beta} = 0. \tag{59b}$$

The commutation relations (57)–(59) are for the full β-algebra, i.e., both spin-0 and spin-1.

Slashed operator products of the type

$$\not{\pi}_n = \prod_{i=1}^{n} \not{a}_i \tag{60}$$

can be reduced by using the above results for products π_n and contracting with the required 4-vectors $(a_i)_\mu$. Alternatively, one can proceed directly from (60) by writing the slashed operators as linear combinations of the basis elements of the respective subalgebras. As in Eqs. (46)–(50), we have

$$\not{a} = \not{a}_R + \not{a}_L. \tag{61}$$

For both spin-0 and spin-1

$$\not{a}_R\not{b}_R = \not{a}_L\not{b}_L = 0, \tag{62}$$

while for spin-0

$$\not{a}_R\not{b}_L = a\cdot bP, \tag{63a}$$

$$\not{a}_L\not{b}_R = \sum_{\lambda\sigma} (a_\lambda b_\sigma)_\lambda P_\sigma, \tag{63b}$$

and for spin-1

$$\not{a}_R\not{b}_L = a\cdot b\sum_\mu {}_\mu V_\mu - \sum_{\lambda\sigma} (a_\lambda b_\sigma)_\sigma V_\lambda, \tag{64a}$$

$$\not{a}_L\not{b}_R = \sum_{\lambda,\sigma,\mu} (a_\lambda b_\sigma)_{\lambda\mu} V_{\mu\sigma}. \tag{64b}$$

With Eqs. (62)–(64) the product $\not{\pi}_n$ can be reduced to linear combinations of the 4-vectors $(a_i)_\mu$ and independent basis elements of the respective subalgebras.

The remaining product of operators which we have to consider is $\not{\beta}^n$. Again we proceed by decomposing the operator $\not{\beta}$ in the form

$$\not{\beta} \equiv \not{\beta}_A + \not{\beta}_B. \tag{65}$$

For spin-0 we define

$$\not{\beta}_A \equiv \sum_\lambda \beta_\lambda(P_\lambda), \qquad \not{\beta}_B \equiv \sum_\lambda \beta_\lambda({}_\lambda P) \tag{66}$$

and for spin-1 we define

$$\not{\beta}_A = \sum_{\lambda,\mu} \beta_\lambda({}_\mu V_{\mu\lambda}), \tag{67a}$$

$$\not{\beta}_B = \sum_{\lambda,\mu} \beta_\lambda({}_{\lambda\mu} V_\mu). \tag{67b}$$

We have defined $\not{\beta}_{A,B}$ in Eqs. (66) and (67) so as to maintain the analogy with the definition of $(\beta_\lambda)_{L,R}$ and $(\not{a})_{L,R}$. However, the similarity is only in the form; $\not{\beta}_A$ $\not{\beta}_B$ are not nilpotent. In fact, for spin-0

$$\not{\beta}_A = \sum_\lambda {}_\lambda P_\lambda, \qquad \not{\beta}_B = 4P \tag{68}$$

and for spin-1

$$\not{\beta}_A = \sum_{\lambda,\mu} {}_{\lambda\mu} V_{\mu\lambda}, \tag{69a}$$

$$\not{\beta}_B = 3\sum_\mu {}_\mu V_\mu. \tag{69b}$$

That is, $\not{\beta}_{A,B}$ are proportional to the basic idempotents of the unit elements for the two subalgebras. For both spin-0 and spin-1

$$\not{\beta}_A\not{\beta}_B = \not{\beta}_B\not{\beta}_A = 0, \tag{70}$$

while for spin-0

$$\not{\beta}_A\not{\beta}_A = \not{\beta}_A, \qquad \not{\beta}_B\not{\beta}_B = 4\not{\beta}_B, \tag{71}$$

and for spin-1

$$\not{\beta}_A\not{\beta}_A = 2\not{\beta}_A, \qquad \not{\beta}_B\not{\beta}_B = 3\not{\beta}_B. \tag{72}$$

To reduce $\not{\beta}^n$ we first use (65) and (70) to write

$$(\not{\beta})^n = (\not{\beta}_A)^n + (\not{\beta}_B)^n. \tag{73}$$

From Eqs. (71) and (72) we have for spin-0

$$(\not{\beta})^n = \not{\beta}_A + 4^{n-1}\not{\beta}_B \tag{74}$$

and for spin-1

$$(\not{\beta})^n = 2^{n-1}\not{\beta}_A + 3^{n-1}\not{\beta}_B. \tag{75}$$

To consider operator products involving combinations of β_λ, $\not{a}$ and $\not{\beta}$ the reduction can always be carried out similarly. In particular, for a product of the form

$$\sigma_{n+m} = \left(\prod_{i=1}^{n} \beta_{\lambda_i}\right)\left(\prod_{j=1}^{m} \not{a}_j\right) \tag{76}$$

the reduction is a straightforward combination of the above results for the products π_n and $\not{\pi}_m$.

For products involving $\not{\beta}$ as well, it follows from Eqs. (58), (59), (74), and (75) that it is sufficient to consider the products $\not{\beta}\pi_n$, $\not{\beta}\not{\pi}_n$, and $\not{\beta}\sigma_n$. To give an example of a reduction of such a product, we will consider $\not{\beta}\pi_n$ explicitly. Writing Eqs. (53) and (54) in the abbreviated form [the superscripts "1" and "2" refer to the first and second terms, respectively, in Eqs. (53) and (54)]

$$\pi_n = \pi_n^1 + \pi_n^2, \tag{77}$$

we have:

Spin-0

$$\not{\beta}\pi_{2n+1} = \beta_A\pi^1_{2n+1} + \beta_B\pi^2_{2n+1} = \pi^1_{2n+1} + 4\pi^2_{2n+1}, \tag{78a}$$

$$\not{\beta}\pi_{2n} = \beta_B\pi^1_{2n} + \beta_A\pi^2_{2n} = 4\pi^1_{2n} + \pi^2_{2n}; \tag{78b}$$

Spin-1

$$\not{\beta}\pi_{2n+1} = \beta_B\pi^1_{2n+1} + \beta_A\pi^2_{2n+1} = 3\pi^1_{2n+1} + 2\pi^2_{2n+1}, \tag{79a}$$

$$\not{\beta}\pi_{2n} = \beta_B\pi^1_{2n} + \beta_A\pi^2_{2n} = 3\pi^1_{2n} + 2\pi^2_{2n}. \tag{79b}$$

To summarize, the above results can be used to reduce operator products involving β_λ, $\not{\pi}$, and $\not{\beta}$ to their subalgebra representations. That is, the reduction gives the product as a linear combination of basis elements of the spin-0 and spin-1 subalgebras.

B. Trace theorems

We now proceed to develop various trace theorems for the β-algebra and the spin-0, -1 subalgebras. We will obtain the desired results by working with the linearly independent basis elements of the P- and V-subalgebras. First we establish two basic theorems for the traces of the subalgebra basis elements.

Theorem 1: For the spin-0 subalgebra the linearly independent basis elements $\{P, P_\mu, {}_\mu P, {}_\mu P_\nu\}$ have the following traces:

$$\mathrm{Tr}P = 1, \tag{80a}$$

$$\mathrm{Tr}({}_\mu P) = \mathrm{Tr}(P_\mu) = 0, \tag{80b}$$

$$\mathrm{Tr}({}_\mu P_\nu) = \delta_{\mu\nu}. \tag{80c}$$

Theorem 2: For the spin-1 subalgebra the linearly independent basis elements $\{{}_\mu V_\nu, {}_\mu V_{\nu\lambda}, {}_{\lambda\nu}V_\mu, {}_{\mu\nu}V_{\lambda\sigma}\}$ have the following traces:

$$\mathrm{Tr}({}_\mu V_\nu) = \delta_{\mu\nu}, \tag{81a}$$

$$\mathrm{Tr}({}_\mu V_{\nu\lambda}) = \mathrm{Tr}({}_{\lambda\nu}V_\mu) = 0, \tag{81b}$$

$$\mathrm{Tr}({}_{\mu\nu}V_{\lambda\sigma}) = \Delta_{\mu\nu\lambda\sigma}. \tag{81c}$$

Theorems 1 and 2 can be established by using the multiplication rules (22) and (38) for the independent basis elements. From Eq. (22) we have that

$$\mathrm{Tr}({}_aP_b)({}_cP_d) = \delta_{bc}\,\mathrm{Tr}({}_aP_d).$$

But

$$\mathrm{Tr}({}_aP_b)({}_cP_d) = \mathrm{Tr}({}_cP_d)({}_aP_b) = \delta_{da}\,\mathrm{Tr}({}_cP_b).$$

Since the set $\{{}_aP_d\}$ is a basis of a complete matrix ring,

$$\mathrm{Tr}({}_aP_d) = \delta_{ad},$$

which establishes Theorem 1. From Eq. (38) we have

$$\mathrm{Tr}({}_{ab}V_{cd})({}_{ef}V_{gh}) = \Delta_{cdef}\,\mathrm{Tr}({}_{ab}V_{gh}).$$

But

$$\mathrm{Tr}({}_{ab}V_{cd})({}_{ef}V_{gh}) = \Delta_{ghab}\,\mathrm{Tr}({}_{ef}V_{cd}).$$

Since the set $\{{}_{ab}V_{gh}\}$ is a basis of a complete matrix ring,

$$\mathrm{Tr}({}_{ab}V_{gh}) = \Delta_{ghab},$$

which establishes Theorem 2.

From Theorems 1 and 2 we immediately have the following corollary results for the traces of independent elements:

Corollary 1: For spin-0 the set of idempotents $\{e_P, P, \sum_\lambda {}_\lambda P_\lambda\}$ have the traces

$$\mathrm{Tr}P = 1; \quad \mathrm{Tr}(\textstyle\sum_\lambda {}_\lambda P_\lambda) = 4, \tag{82a}$$

$$\mathrm{Tr}\,e_P = 5. \tag{82b}$$

Corollary 2: For spin-1 the set of idempotents $\{e_V, \sum_\mu {}_\mu V_\mu, \frac{1}{2}\sum_{\mu,\nu} {}_{\mu\nu}V_{\nu\mu}\}$ have the following traces:

$$\mathrm{Tr}(\textstyle\sum_\mu {}_\mu V_\mu) = 4, \quad \mathrm{Tr}(\frac{1}{2}\sum_{\mu,\nu} {}_{\mu\nu}V_{\nu\mu}) = 6, \tag{83a}$$

$$\mathrm{Tr}\,e_V = 10. \tag{83b}$$

Using Theorems 1 and 2 in conjunction with Eqs. (53) and (54) the following three theorems are readily proven.

Theorem 3: For both spin-0 and spin-1

$$\mathrm{Tr}(\pi_{2n+1}) = 0, \tag{84a}$$

where

$$\pi_{2n+1} = \prod_{i=1}^{2n+1} \beta_{\lambda_i}. \tag{84b}$$

Theorem 4: For the spin-0 subalgebra the trace of the product of an even number of β matrices is

$$\mathrm{Tr}(\pi_{2n}) = \prod_{i=1}^{n} \delta(\lambda_{2i-1}, \lambda_{2i}) + \delta(\lambda_1, \lambda_{2n}) \left(\prod_{i=1}^{n-1} \delta(\lambda_{2i}, \lambda_{2i+1})\right), \tag{85a}$$

where

$$\pi_{2n} \equiv \prod_{i=1}^{2n} \beta_{\lambda_i}. \tag{85b}$$

Theorem 5: For the spin-1 subalgebra the trace of the product of an even number of β matrices is given by

$$\mathrm{Tr}\pi_2 = 2\sum_{\mu_1} \Delta(\mu_1, \lambda_1, \lambda_2, \mu_1) = 6\delta(\lambda_1, \lambda_2), \tag{86a}$$

and for $n > 1$

$$\mathrm{Tr}\pi_{2n} = \sum_{\mu} \Delta(\mu_{2n}, \lambda_1, \lambda_2, \mu_2) \left(\prod_{i=2}^{n} \Delta(\mu_{2i-2}, \lambda_{2i-1}, \lambda_{2i}, \mu_{2i}) \right) + \sum_{\mu} \Delta(\lambda_1, \mu_2, \lambda_{2n}, \mu_{2n}) \left(\prod_{i=2}^{n} \Delta(\mu_{2i-2}, \lambda_{2i-2}, \lambda_{2i-1}, \mu_{2i}) \right) \tag{86b}$$

where in the above the sum over μ means all μ_i that occur in the equation and

$$\pi_{2n} = \prod_{i=1}^{2n} \beta_{\lambda_i} \quad (n = 2, \ldots). \tag{86c}$$

The following theorem can be obtained from Corollaries 1 and 2 and the results contained in Eqs. (65)–(75).

Theorem 6: For the operator $\not{\beta}$ we have the following traces:

Spin-0

$$\mathrm{Tr}\not{\beta}_A = 4, \quad \mathrm{Tr}\not{\beta}_B = 4, \quad \mathrm{Tr}(\not{\beta})^n = 4 + 4^n, \quad n = 1, 2, \ldots, \tag{87}$$

Spin-1

$$\mathrm{Tr}\not{\beta}_A = 12, \quad \mathrm{Tr}\not{\beta}_B = 12, \quad \mathrm{Tr}(\not{\beta})^n = 12(2^{n-1} + 3^{n-1}), \quad n = 1, 2, \ldots. \tag{88}$$

An important set of products for which it is desirable to know the traces is

$$\not{\pi}_n = \prod_{i=1}^{n} \not{\phi}_i$$

and

$$\sigma_{n+m} = \left(\prod_{i=1}^{n} \not{\phi}_i \right) \left(\prod_{j=1}^{m} \beta_{\lambda_j} \right)$$

or

$$\sigma_{n+m} = \left(\prod_{j=1}^{m} \beta_{\lambda_j} \right) \left(\prod_{i=1}^{n} \not{\phi}_i \right).$$

For the cases where the above products contain an odd number of β matrices we have the following corollary to Theorem 3.

Corollary 3: For both spin-0 and spin-1 subalgebras

$$\mathrm{Tr}\not{\pi}_{2n+1} = \mathrm{Tr}\sigma_{2n+2m+1} = 0. \tag{89}$$

For the case of $\not{\pi}_{2n}$ the trace of an even numbered product of slashed operators is given by the following corollary:

Corollary 4: For both spin-0 and spin-1 subalgebras the trace of $\not{\pi}_{2n}$ is given by

$$\mathrm{Tr}\not{\pi}_{2n} = \mathrm{Tr}\left(\prod_{i=1}^{2n} \beta_{\lambda_i} \right) \left(\prod_{i=1}^{2n} (a_i)_{\lambda_i} \right) \equiv \mathrm{Tr}(\pi_{2n}) \left(\prod_{i=1}^{2n} (a_i)_{\lambda_i} \right). \tag{90}$$

The remaining result to consider is for operator products involving $\not{\beta}$. By Eqs. (74) and (75) it is sufficient to consider $\not{\beta}\pi_{2n}$ and $\not{\beta}\not{\pi}_{2n}$. The required theorem then follows directly from Eqs. (78) and (79).

Theorem 7: The trace of $\not{\beta}\pi_{2n}$ for the case of spin-0 is

$$\mathrm{Tr}\not{\beta}\pi_{2n} = 4\,\mathrm{Tr}\pi^1_{2n} + \mathrm{Tr}\pi^2_{2n}, \tag{91a}$$

and for spin-1 is

$$\mathrm{Tr}\not{\beta}\pi_{2n} = 3\,\mathrm{Tr}\pi^1_{2n} + 2\,\mathrm{Tr}\pi^2_{2n}. \tag{91b}$$

Finally, we collect together various trace results under a title of a final theorem.

Theorem 8: For the spin-0 subalgebra there are the following miscellaneous trace results:

$$\mathrm{Tr}\beta_\mu^2 = 2 \quad \text{(no summation!)}, \tag{92a}$$

$$\mathrm{Tr}\eta_\mu = -1, \tag{92b}$$

$$\mathrm{Tr}\xi^+ = 2, \tag{92c}$$

$$\mathrm{Tr}\xi^- = 3. \tag{92d}$$

For the spin-1 subalgebra the corresponding trace results are

$$\mathrm{Tr}\beta_\mu^2 = 6 \quad \text{(no summation!)}, \tag{93a}$$

$$\mathrm{Tr}\eta_\mu = 2, \tag{93b}$$

$$\mathrm{Tr}\xi^+ = 6, \tag{93c}$$

$$\mathrm{Tr}\xi^- = 4. \tag{93d}$$

5. PROJECTION, RAISING AND LOWERING OPERATORS, AND THE DKP CONSEQUENT EQUATIONS

In this section we will first show that the P- and V-subalgebra basis elements are complete sets of projection, raising and lowering operators in the space of the DKP multiple field components. This will then aid us in examining the DKP consequent equations in the irreducible representations for spin-0 and spin-1.

A. Operators

The interpretation of the P- and V-subalgebra basis elements as complete sets of projection, raising and lowering operators follows directly by noting that the subalgebra basis elements have Kronecker product representations.

For spin-0 we take the basis elements

$$\{K(5)\} = \{{}_aP_b \mid a, b = 1, \ldots, 5\}, \tag{94}$$

which have the Kronecker representation

$$({}_aP_b)_{cd} = \delta_{ca}\delta_{db}. \tag{95}$$

(See Appendix 2 for the Pauli metric representation of the P-subalgebra.)

If we now ask what new five-component field is obtained after operating with one of the basis elements, we have

$$\psi' = {}_aP_b\psi, \tag{96a}$$

$$\psi'_c = \delta_{ca}\psi_b. \tag{96b}$$

Thus we find that the five elements ${}_aP_a$ project from ψ the ath components. For $a < b$ the elements ${}_aP_b$ project from ψ the bth component and raise it to the ath position. Likewise the elements ${}_aP_b$ $(a > b)$ are lowering operators. It might be noted as well that $\sum_{a=1}^{4} {}_aP_a$ projects from the field the four components that transform

as a 4-vector. The element $P \equiv {}_5P_5$ projects the component that transforms as a scalar.

For spin-1 we have the independent basis elements

$$\{K(10)\} = \{{}_{ab}V_{cd} \mid a, b, c, d = 1, \ldots, 5\}. \tag{97}$$

with

$${}_{ab}V_{cd} = -{}_{ab}V_{dc} = {}_{ba}V_{dc}. \tag{98}$$

From Appendix 2 with $\{m, n = 1, \ldots, 10\}$ and $[cd]$ the ordered set $[12] \sim 1$, $[13] \sim 2$, $[14] \sim 3$, $[15] \sim 4$, $[23] \sim 5$, $[24] \sim 6$, $[25] \sim 7$, $[34] \sim 8$, $[35] \sim 9$, $[45] \sim 10$, the Kronecker product representation of the basis element is

$$({}_{ab}V_{cd})_{mn} = \delta'(m, [ba])\delta'(n, [cd]), \tag{99}$$

where

$$\delta'(n, [cd]) = \begin{cases} \delta(n, [cd]), & c < d \quad (100a) \\ -\delta(n, [dc]), & c > d. \quad (100b) \end{cases}$$

As for the case of spin-0, the new field obtained by operating with the basis elements is

$$\psi' = {}_{ab}V_{cd}\psi, \quad a > b, \quad d > c, \tag{101a}$$

$$\psi'_m = \delta'(m, [ba])\psi_{[cd]}. \tag{101b}$$

Again we have that for $[ba] \equiv [cd]$, ${}_{ab}V_{cd}$ projects from ψ the $[cd]$ component and for $[ba] < [cd]$ or $[ba] > [cd]$, ${}_{ab}V_{cd}$ is a raising or lowering operator, respectively.

B. Consequent equations

We now examine the DKP "consequent equations" using the irreducible representations for spin-0 and spin-1. In particular, we will show that solutions of the consequent equations are given by solutions of Eqs. (13) and (16). The DKP consequent equations are[2]

$$\partial_\lambda\psi = \partial_\rho\beta_\rho\beta_\lambda\psi. \tag{102}$$

We write these equations for spin-0 as

$$\partial_\lambda\psi = (\partial_\lambda)P\psi + (\partial_\rho)_\rho P_\lambda\psi, \tag{103}$$

and for spin-1 as

$$\partial_\lambda\psi = \partial_\lambda(\textstyle\sum_\mu {}_\mu V_\mu)\psi - \partial_\rho({}_\lambda V_\rho)\psi + \partial_\rho(\sum_\mu {}_{\rho\mu}V_{\mu\lambda})\psi. \tag{104}$$

Using the unit matrix in the two representations, Eqs. (103) and (104) reduce for spin-0 to

$$\partial_\lambda(\textstyle\sum_\mu {}_\mu P_\mu)\psi = \partial_\rho({}_\rho P_\lambda)\psi \tag{105}$$

and for spin-1 to

$$\partial_\lambda\left(\frac{1}{2}\sum_{\mu\nu} {}_{\mu\nu}V_{\nu\mu}\right)\psi = -\partial_\rho({}_\lambda V_\rho)\psi + \partial_\rho\sum_\mu({}_{\rho\mu}V_{\mu\lambda})\psi. \tag{106}$$

Recall the just derived properties of the basis elements being projection and raising and lowering operators. Using these properties it can be seen that Eq. (105) is an equation only for the four components of the spin-0 field ψ that transform as a 4-vector.

However, the lhs and the second term on the rhs of Eq. (106) are zero for the four components of the spin-1 field that transform as a 4-vector. Further, the first term on the rhs of Eq. (106) is zero for the other six components. These two statements combine to mean that the spin-1 Eq. (106) implies

$$\partial_\rho({}_\lambda V_\rho)\psi = 0. \tag{107}$$

Equation (107) represents the subsidiary condition on the field to eliminate timelike mesons. In terms of the R-algebra we note that

$${}_4V_\rho \equiv R_\rho. \tag{108}$$

Thus Eq. (107) can be written

$$\partial_\rho(R_\rho)\psi = 0. \tag{109}$$

Obtaining the constraint on the remaining field components (one remaining for spin-0, six remaining for spin-1) implied by Eqs. (105) and (106) is most easily done by considering the consequent Eq. (102) in an alternative form. Equation (102) is derived by combining the equation

$$\partial_\rho\beta_\rho\beta_\sigma(\beta_\mu\partial_\mu + m)\psi = 0 \tag{110}$$

and the defining relation (2) for the β matrices.

Starting from (110) and using the P-algebra representation for spin-0, we have

$$\partial_\sigma\partial_\mu(P_\mu)\psi + \partial_\rho\partial_\sigma({}_\rho P)\psi = -m\partial_\sigma(P)\psi - m\partial_\rho({}_\rho P_\sigma)\psi. \tag{111}$$

Multiplying Eq. (111) on the left by P we obtain

$$\partial_\mu(P_\mu)\psi = -mP\psi. \tag{112a}$$

Likewise, multiplying Eq. (111) on the left by P_ν, we obtain

$$\partial_\sigma(P)\psi = -mP_\sigma\psi. \tag{112b}$$

Equations (112a) and (112b) are identical to Eq. (13).

Similarly, after somewhat tedious algebra, it can be shown that the consequent equations for spin-1 have the solution (16).

6. APPLICATIONS TO SPIN-0 MESONS

Recently[11-13] we have initiated a program to analyze the interactions of spin-0 mesons using the DKP equation instead of the Klein-Gordon equation. The analysis has led to an improved parameterization of the mesonic matrix elements. It is the purpose of this section to apply the algebraic formalism introduced in the previous section to show how it facilitates the analysis of such problems. By way of example, it will be shown that the P-algebra is a natural formalism to use to compare DKP results with results based on the Klein-Gordon equation. We will first consider the trivial case of free particles and then consider cases of interacting particles.

A. Free field case

The Klein-Gordon equations for a charged spin-0 particle are

$$\partial_\mu\partial_\mu\phi = m^2\phi, \qquad \partial_\mu\partial_\mu\phi^* = m^2\phi^*. \tag{113}$$

The conserved charge current is

$$j_\mu^{KG} = -i\phi^* \overleftrightarrow{\partial}_\mu \phi = -i[\phi^* \partial_\mu \phi - (\partial_\mu \phi^*)\phi]. \tag{114}$$

The normalized plane wave solutions of (113) are

$$\phi(\mathbf{x}, t) = (1/\sqrt{2p_0 V})e^{ip\cdot x} \tag{115}$$

$$\phi^*(\mathbf{x}, t) = (1/\sqrt{2p_0 V})e^{-ip\cdot x}.$$

The DKP equations are

$$(\partial_\mu \beta_\mu + m)\psi = 0, \quad \partial_\mu \bar{\psi}\beta_\mu - m\bar{\psi} = 0, \tag{116a}$$

where

$$\bar{\psi} \equiv \psi^\dagger \eta_4 \tag{116b}$$

and the conserved current associated with the particle is

$$j_\mu^{DKP} = i\bar{\psi}\beta_\mu \psi. \tag{117}$$

Normalized plane wave solutions of the DKP equations for spin-0 particles are

$$\psi(\mathbf{x}, t) = (m/p_0 V)^{1/2} u(p) e^{ip\cdot x}, \qquad \bar{\psi}(\mathbf{x}, t) = (m/p_0 V)^{1/2} \bar{u}(p) e^{-ip\cdot x} \tag{118}$$

where the five-component wave functions are

$$u(p) = \frac{1}{\sqrt{2m^2}} \begin{pmatrix} ip_0 \\ ip_x \\ ip_y \\ ip_z \\ -m \end{pmatrix}, \tag{119a}$$

$$\bar{u}(p) = (1/\sqrt{2m^2})(-ip_0, ip_x, ip_y, ip_z, -m). \tag{119b}$$

The β-algebra as defined by Eq. (2) imposes constraints on the DKP field, namely,

$$\partial_\mu \psi = \partial_\nu \beta_\nu \beta_\mu \psi, \tag{120a}$$

$$\partial_\mu \bar{\psi} = \partial_\nu \bar{\psi} \beta_\mu \beta_\nu. \tag{120b}$$

These constraint equations will be useful in later discussions.

Reducing the β-algebra to the spin-0 subalgebra and using the P-algebra representation, the DKP equations read

$$\partial_\mu(P)\psi = -m(P_\mu)\psi, \qquad \partial_\mu(P_\mu)\psi = -mP\psi \tag{121}$$

and

$$\partial_\mu \bar{\psi} P = m\bar{\psi}_\mu P, \qquad \partial_\mu \bar{\psi}_\mu P = m\bar{\psi}P. \tag{122}$$

The constraint equations (120) become

$$\partial_\mu \psi = \partial_\mu(P)\psi + \partial_\nu({}_\nu P_\mu)\psi, \qquad \partial_\mu \bar{\psi} = \partial_\mu \bar{\psi} P + \partial_\nu \bar{\psi}({}_\mu P_\nu). \tag{123}$$

Now consider the free particle current in the DKP formalism. For spin-0 the current is

$$j_\mu^{DKP} = i\bar{\psi}\beta_\mu\psi = i\bar{\psi}(P_\mu + {}_\mu P)\psi. \tag{124}$$

From Eqs. (121) and (122) we can write (124) as

$$j_\mu^{DKP} = i\bar{\psi}(P_\mu)\psi + i\bar{\psi}({}_\mu P)\psi = -\frac{i}{m}(\bar{\psi}P)\partial_\mu(P\psi) + \frac{i}{m}(\partial_\mu \bar{\psi}P)(P\psi). \tag{125}$$

Since $P\psi$ and $\bar{\psi}P$ are solutions of the KG equation and are just the fifth components of the DKP fields, the KG and DKP free particle currents are identical.

B. Conserved interaction current

In the case of interacting particles the simplest case we can consider is the interaction of spin-0 particles with an electromagnetic field. Introducing the electromagnetic field by the minimal substitution

$$\partial_\mu \to \partial_\mu \mp ieA_\mu = \partial_\mu^\mp, \tag{126}$$

the Klein–Gordon equations become

$$\partial_\mu^- \partial_\mu^- \phi = m^2\phi, \qquad \partial_\mu^+ \partial_\mu^+ \phi^* = m^2\phi^*. \tag{127}$$

The interacting meson current is

$$j_\mu^{KG} = -i[\phi^* \partial_\mu^- \phi - (\partial_\mu^+ \phi^*)\phi], \tag{128}$$

which is conserved:

$$\partial_\mu j_\mu^{KG} = 0. \tag{129}$$

With minimal substitution in the DKP equations the interacting field equations are

$$(\partial_\mu^- \beta_\mu + m)\psi = 0, \qquad \partial_\mu^+ \bar{\psi}\beta_\mu - m\bar{\psi} = 0. \tag{130}$$

The conserved current is

$$j_\mu^{DKP} = i\bar{\psi}\beta_\mu\psi, \qquad \partial_\mu j_\mu^{DKP} = 0. \tag{131}$$

The field equations that are a consequence of the β-algebra in this case are

$$\partial_\mu^- \psi = \partial_\nu^- \beta_\nu \beta_\mu \psi + \frac{ie}{2m} F_{\nu\rho}(\beta_\rho \beta_\mu \beta_\nu - \delta_{\rho\mu}\beta_\nu)\psi, \qquad \partial_\mu^+ \bar{\psi} = \partial_\nu^+ \bar{\psi}\beta_\mu\beta_\nu + \frac{ie}{2m} F_{\nu\rho}\bar{\psi}(\beta_\nu\beta_\mu\beta_\rho - \delta_{\rho\mu}\beta_\nu), \tag{132}$$

where we have introduced the electromagnetic field strengths via

$$\pm ieF_{\nu\rho} = \partial_\nu^\pm \partial_\rho^\pm - \partial_\rho^\pm \partial_\nu^\pm = \pm ie(\partial_\nu A_\rho - \partial_\rho A_\nu). \tag{133}$$

Going over to the second order wave equations from (132), we have

$$\partial_\mu^- \partial_\mu^- \psi = m^2\psi + ieF_{\mu\nu}\beta_\mu\beta_\nu\psi + \frac{ie}{2m}\partial_\mu^- F_{\nu\rho}(\beta_\rho\beta_\mu\beta_\nu - \delta_{\rho\mu}\beta_\nu)\psi,$$

$$\partial_\mu^+ \partial_\mu^+ \bar{\psi} = m^2\bar{\psi} - ieF_{\mu\nu}\bar{\psi}\beta_\mu\beta_\nu + \frac{ie}{2m}\partial_\mu^+ F_{\nu\rho}\bar{\psi}(\beta_\nu\beta_\mu\beta_\rho - \beta_\nu\delta_{\rho\mu}). \tag{134}$$

Finally, we have that the current can be decomposed as in the case of a Dirac particle so that

$$j_\mu^{\mathrm{DKP}} = i\bar{\psi}\beta_\mu\psi = \frac{i}{2m}\{(\partial_\mu^+\bar{\psi})\,\psi - \bar{\psi}\partial_\mu^-\psi + \partial_\nu[\bar{\psi}(\beta_\nu\beta_\mu - \beta_\mu\beta_\nu)\psi] - \frac{ie}{m}F_{\nu\rho}\bar{\psi}\beta_\nu\beta_\mu\beta_\rho\psi\}. \tag{135}$$

(Note that at this point our discussion of a DKP particle interacting with an electromagnetic field holds for either spin-0 or spin-1 since we have used the reducible β-algebra.)

We should remark at this point that upon comparing Eqs. (127) and (128) with Eqs. (134) and (135) one might be lead to suspect that the conserved electrodynamics of spin-0 particles in the Klein-Gordon and DKP formalisms differ. Such a difference might arise since the interaction is introduced by the same assumption in different equations. However, we will in fact prove that there is no difference. First let us examine the structure of the additional DKP terms in the spin-0 case by introducing the P-algebra.

The interacting field Eq. (130) becomes

$$\partial_\mu^- P\psi = -mP_\mu\psi, \qquad \partial_\mu^- P_\mu\psi = -mP\psi \tag{136}$$

and

$$\partial_\mu^+\bar{\psi}P = m\bar{\psi}_\mu P, \qquad \partial_\mu^+\bar{\psi}_\mu P = m\bar{\psi}P. \tag{137}$$

In terms of the P-algebra the consequent Eqs. (132) and the second order Eqs. (134) are

$$\partial_\mu^-\psi = \partial_\mu^- P\psi + \partial_\nu^-({}_\nu P_\mu)\psi + \frac{ie}{m}F_{\mu\nu}({}_\nu P)\psi, \qquad \partial_\mu^+\bar{\psi} = \partial_\nu^+\bar{\psi}P + \partial_\nu^+\bar{\psi}({}_\mu P_\nu) + \frac{ie}{m}F_{\mu\nu}\bar{\psi}P_\nu \tag{138}$$

and

$$\partial_\mu^-\partial_\mu^-\psi = m^2\psi + ieF_{\mu\nu}({}_\mu P_\nu)\psi + \frac{ie}{m}\partial_\mu^- F_{\mu\nu}({}_\nu P)\psi, \qquad \partial_\mu^+\partial_\mu^+\bar{\psi} = m^2\bar{\psi} - ieF_{\mu\nu}\bar{\psi}({}_\nu P_\mu) + \frac{ie}{m}\partial_\mu^+ F_{\mu\nu}\bar{\psi}P_\nu. \tag{139}$$

The expression (68) for the current becomes

$$j_\mu^{\mathrm{DKP}} = \frac{i}{2m}\left\{(\partial_\mu^+\bar{\psi})\psi - \bar{\psi}(\partial_\mu^-\psi) + \partial_\nu[\psi({}_\nu P_\mu - {}_\mu P_\nu)\psi] - \frac{ie}{m}F_{\mu\nu}\bar{\psi}(P_\nu - {}_\nu P)\psi\right\}. \tag{140}$$

Although we have specialized to the case of spin-0 via the P-algebra, the additional interaction terms are still formally present in the DKP Eqs. (139) and (140). The appearance of these additional terms arises from the vector component of the five-component DKP field. This can be seen from the fact that

$$\partial_\mu^-\partial_\mu^-(P\psi) = m^2(P\psi), \qquad \partial_\mu^+\partial_\mu^+(\bar{\psi}P) = m^2(\bar{\psi}P), \tag{141}$$

which follow from either Eqs. (136) and (137) or from Eqs. (139). The second order equations for the vector components of the DKP fields follows from (139) and are

$$\partial_\mu^-\partial_\mu^-({}_\lambda P)\psi = m^2({}_\lambda P)\psi + 2ieF_{\lambda\mu}({}_\mu P)\psi, \qquad \partial_\mu^+\partial_\mu^+\bar{\psi}({}_\lambda P) = m^2\bar{\psi}({}_\lambda P) - 2ieF_{\lambda\mu}\bar{\psi}({}_\mu P). \tag{142}$$

We will now proceed to prove that these apparently additional interaction terms are only part of the formalism, and the interaction in the DKP and Klein-Gordon formalisms are indeed the same.

The proof is simple. Rather than writing the current decomposition (140) one proceeds directly from the expression for the current before decomposition. We had

$$j_\mu^{\mathrm{DKP}} = i\bar{\psi}\beta_\mu\psi = i\bar{\psi}(P_\mu - {}_\mu P)\psi. \tag{143}$$

From (136) and (137) we then have

$$j_\mu^{\mathrm{DKP}} = -\frac{i}{m}\bar{\psi}\partial_\mu P\psi + \frac{i}{m}(\partial_\mu^+\bar{\psi}P) = -\frac{i}{m}[(\bar{\psi}P)\partial_\mu^-(P\psi) - (\partial_\mu^+\bar{\psi}P)(P\psi)]. \tag{144}$$

But since $\bar{\psi}P$ and $P\psi$ are solutions of the KG Eqs. (141) the two currents (128) and (144) are identical.

C. Nonconserved electrodynamics

It should be emphasized that in the above proof the equivalence of the Klein-Gordon and DKP descriptions of interacting spin-0 particles is only true for a conserved interaction charge. To see what happens when the interaction charge is not conserved let us remove the restriction that the electric charge be conserved. Retaining the minimal substitution prescription for introducing the electromagnetic field, we have the partially conserved currents

$$j_\mu^{\mathrm{KG}} = -i[\phi_B^*\partial_\mu^-\phi_A - (\partial_\mu^+\phi_B^*)\phi_A], \qquad \partial_\mu j_\mu^{\mathrm{KG}} = -i(m_A^2 - m_B^2)\phi_B^*\phi_A \tag{145}$$

and

$$j_\mu^{\mathrm{DKP}} = i\bar{\psi}_B\beta_\mu\psi_A, \qquad \partial_\mu j_\mu^{\mathrm{DKP}} = i(m_B - m_A)\bar{\psi}_B\psi_A. \tag{146}$$

In the limit of the current being conserved we have $m_A = m_B$. It will be seen below that the mass difference associated with a partially conserved charge leads to a difference between the two formalisms.

Proceeding as in the case of a conserved current, we have

$$j_\mu^{\mathrm{DKP}} = \frac{-i}{m_A}(\bar{\psi}_B P)\partial_\mu^-(P\psi_A) + \frac{i}{m_B}(\partial_\mu^+\bar{\psi}_B P)(P\psi_A). \tag{147}$$

Now using the relations

$$(P\psi_A) \equiv -\sqrt{m_A}\,\phi_A, \qquad (\bar{\psi}_B P) \equiv -\sqrt{m_B}\,\phi_B^*, \tag{148}$$

Eq. (147) becomes

$$j_\mu^{\mathrm{DKP}} = -i\sqrt{m_B/m_A}\,\phi_B^*\partial_\mu^-\phi_A + i\sqrt{m_A/m_B}(\partial_\mu^+\phi_B^*)\phi_A, \tag{149}$$

which is clearly not Eq. (145) for the Klein-Gordon current.[14] Note, however, that again in the limit $m_A = m_B$ the two currents become identical.

Equation (149) can be written in the more interesting form

$$j_\mu^{\mathrm{DKP}} = -\frac{i(m_B+m_A)}{2\sqrt{m_A m_B}}\left(\phi_B^* \partial_\mu^- \phi_A - (\partial_\mu^+ \phi_B^*)\phi_A\right) - \frac{i(m_B-m_A)}{2\sqrt{m_A m_B}}\left(\phi_B^* \partial_\mu^- \phi_A + (\partial_\mu^+ \phi_B^*)\phi_A\right). \quad (150)$$

Equation (150) consists of two pieces. The first piece is the Klein–Gordon current multiplied by a factor which is unity only when $m_A = m_B$. The second piece does not occur in the Klein–Gordon current and would have to be included via an induced coupling term. Thus we can conclude that using the same prescription for introducing an interaction in the DKP and KG formalisms can lead to inequivalent results for the case of a partially conserved current.

At first sight, this conclusion appears to be surprising. However, it should be noted that although there is an intimate connection between induced couplings in the case of a partially conserved current and the mass difference, the connection is not as obvious as is often assumed.

D. General nonconserved currents

It is relevant to comment here that not only does the DKP current have a different structure but the current divergence in (146) may have a zero at $t = [(m_A + m_B)^2]$.[11,15] It is obvious that the two additional features are related. A detailed examination of the DKP and KG descriptions of partially conserved interaction charge currents is presented elsewhere.[14] The question of the formal structure of these currents is in keeping with our present discussion and it is to this question that we will now address ourselves.

In some earlier work[11–13,16] we have been particularly interested in the nonconserved vector-current matrix element between pseudoscalar (scalar) meson states. (If the states were of opposite intrinsic parity, the relevant matrix element would be the axial current matrix element and all of the subsequent arguments for the case of a vector current would apply.) Since these are partially conserved currents, we should also consider the current-divergence matrix element. Indeed, since the current-divergence matrix element has a simpler structure, we will consider it first.

For the case of a KG description we have

$$\langle B|\partial_\lambda V_\lambda^{\mathrm{KG}}(0)|A\rangle = \frac{i(m_B^2 - m_A^2)}{2V\sqrt{E_A E_B}} f_0(t), \quad (151)$$

where

$$t = -(p_A - p_B)^2.$$

In the limit that the masses of the two states become equal the current-divergence is zero. For the DKP description of the current-divergence there are two possible terms:

$$D_a = \partial_\lambda V_\lambda^a = \bar\psi_B P \psi_A g_a(t), \qquad D_b = \partial_\lambda V_\lambda^b = \bar\psi_B(\textstyle\sum_\lambda P_\lambda)\psi_A g_b(t). \quad (152)$$

However, as one would expect, the two possible terms are not independent since

$$\bar\psi_B(\textstyle\sum_\lambda P_\lambda)\psi_A = -\frac{p_B \cdot p_A}{m_A m_B}\bar\psi_B P \psi_A. \quad (153)$$

Thus we have

$$D_b = \bar\psi_B P \psi_A g_b'(t). \quad (154)$$

The particular choice for the current-divergence in the DKP formalism must be made on an independent basis, such as dynamical considerations. In earlier work we were lead to make the choice

$$D = \partial_\lambda V_\lambda = i(m_B - m_A)\bar\psi_B \psi_A g_0(t) = i(m_B - m_A)\bar\psi_B(P + \textstyle\sum_\lambda P_\lambda)\psi_A g_0(t). \quad (155)$$

This choice is "natural" in the sense that for physical DKP fields in the absence of explicit induced coupling we have

$$\partial_\lambda V_\lambda = i(m_B - m_A)\bar\psi_B \psi_A.$$

A point that should be emphasized here concerns the question of so-called "smoothness" assumptions for form factors such as $g_0(t)$. This question arises because we can write (155) as

$$\langle B|D|A\rangle = \frac{i(m_B^2 - m_A^2)}{2V\sqrt{E_A E_B}} f_0(t) g_0(t), \qquad f_0(t) \equiv \left((m_A + m_B)^2 - t\right)/2(m_A + m_B)(m_A m_B)^{1/2}. \quad (156)$$

The $f_0(t)$ t-dependence comes from the wave functions. Smoothness assumptions about induced couplings apply to $g_0(t)$. The choice for $f_0(t)$ is an assumption about the wave functions for the physical states and, in particular, is an assumption about the dependence of the current-divergence on the mass difference.

We turn now to a discussion of the vector current matrix element. In the KG formalism this is

$$\langle B|V_\lambda^{\mathrm{KG}}(0)|A\rangle = \frac{1}{2V\sqrt{E_A E_B}}[(p_A + p_B)_\lambda f_+(t) + (p_A - p_B)_\lambda f_-(t)]. \quad (157)$$

For spin-0 particles the only 4-vectors in the problem are the momenta of the particles and consequently the parameterization of Eq. (157) is necessary in the KG description. In the DKP formalism there are additional 4-vectors arising from the β-algebra. However, as one would expect, there can be only two independent form factors in the parameterization of a vector current matrix element. For completeness, and to demonstrate the utility of our algebraic formalism, we will explicitly prove that there are only two independent 4-vectors. Moreover, a detailed examination of possible induced vector couplings is important for the reasons raised in the above discussion of the current-divergence matrix element.

The linearly independent basis elements of the spin-0 subalgebra are

$$\{P, P_\lambda, {}_\lambda P, {}_\lambda P_\nu\}.$$

For the present discussion it is convenient to transform to an equivalent set of linearly independent basis elements, namely,

$$\{(P), (P_\lambda + {}_\lambda P), (P_\lambda - {}_\lambda P), ({}_\lambda P_\nu + {}_\nu P_\lambda), ({}_\lambda P_\nu - {}_\nu P_\lambda)\}.$$

In terms of the five-dimensional representation of the β-algebra these elements are

$$P = \beta_\mu^2 \beta_\nu^2, \quad \mu \neq \nu,$$

$$P_\lambda + {}_\lambda P = \beta_\lambda,$$

$$P_\lambda - {}_\lambda P = \frac{1}{3}\sum_\nu(\beta_\nu^2 \beta_\lambda - \beta_\lambda \beta_\nu^2),$$

$$
\begin{aligned}
{}_{\lambda}P_{\nu} + {}_{\nu}P_{\lambda} &= \beta_{\lambda}\beta_{\nu} + \beta_{\nu}\beta_{\lambda}, \quad \nu \neq \lambda \\
&= \beta_{\lambda}^2(1-\beta_{\mu}^2), \quad \nu = \lambda \neq \mu, \\
{}_{\lambda}P_{\nu} - {}_{\nu}P_{\lambda} &= \beta_{\lambda}\beta_{\nu} - \beta_{\nu}\beta_{\lambda}.
\end{aligned}
$$

The possible vector couplings, including induced derivative couplings, are

$$V_{\lambda}^{1} = \bar{\psi}_B(P_{\lambda} + {}_{\lambda}P)\psi_A g_1(t), \tag{158a}$$

$$V_{\lambda}^{2} = \bar{\psi}_B(P_{\lambda} - {}_{\lambda}P)\psi_A g_2(t), \tag{158b}$$

$$V_{\lambda}^{3} = \partial_{\lambda}(\bar{\psi}_B P\psi_A)g_3(t), \tag{158c}$$

$$V_{\lambda}^{4} = \partial_{\nu}[\bar{\psi}_B({}_{\lambda}P_{\nu} + {}_{\nu}P_{\lambda})\psi_A]g_4(t), \tag{158d}$$

$$V_{\lambda}^{5} = \partial_{\nu}[\bar{\psi}_B({}_{\lambda}P_{\nu} - {}_{\nu}P_{\lambda})\psi_A]g_5(t), \tag{158e}$$

$$V_{\lambda}^{6} = \partial_{\lambda}(\bar{\psi}_B\psi_A)g_6(t). \tag{158f}$$

Using Eqs. (121) and (122), the current element (158c) becomes

$$
\begin{aligned}
V_{\lambda}^{3} = \frac{g_3(t)}{2m_A m_B}[&(m_A - m_B)\bar{\psi}_B(P_{\lambda} + {}_{\lambda}P)\psi_A \\
&- (m_A + m_B)\bar{\psi}_B(P_{\lambda} - {}_{\lambda}P)\psi_A].
\end{aligned} \tag{159}
$$

Using the results from the discussion of current-divergence the current element (158f) can be reduced to

$$V_{\lambda}^{6} = \partial_{\lambda}(\bar{\psi}_B P\psi_A)g_6'(t). \tag{160}$$

To reduce the current elements (158d) and (158e) the following two identities are required:

$$(\partial_{\nu}\bar{\psi}_B)({}_{\mu}P_{\nu})\psi_A + \bar{\psi}_B({}_{\nu}P_{\mu})\partial_{\nu}\psi_A = \partial_{\mu}(\bar{\psi}_B\psi_A) - \partial_{\mu}(\bar{\psi}_B P\psi_A) \tag{161}$$

and

$$
\begin{aligned}
(\partial_{\nu}\bar{\psi}_B)({}_{\mu}P_{\nu})\psi_A &- \bar{\psi}_B({}_{\nu}P_{\mu})\partial_{\nu}\psi_A \\
&= \left(\frac{(m_A + m_B)}{2m_A m_B}\bar{\psi}_B(P_{\mu} + {}_{\mu}P)\psi_A\right. \\
&\quad \left. + \frac{(m_B - m_A)}{2m_A m_B}\psi_B(P_{\mu} - {}_{\mu}P)\psi_A\right) h_1(t).
\end{aligned} \tag{162}
$$

(All $h_i(t)$ are explicit functions of t.) The identities (161) and (162) can be deduced from the constraint equations (123). Using these identities the current elements (158d) and (158e) become

$$
\begin{aligned}
V_{\lambda}^{4} = &\left(\frac{(m_A + m_B)}{2m_A m_B}\bar{\psi}_B(P_{\lambda} - {}_{\lambda}P)\psi_A\right. \\
&\left. + \frac{(m_A - m_B)}{2m_A m_B}\bar{\psi}_B(P_{\lambda} + {}_{\lambda}P)\psi\right) g_4(t) \\
&+ \partial_{\lambda}(\bar{\psi}_B\psi_A) - \partial_{\lambda}(\bar{\psi}_B P\psi_A)g_4(t)
\end{aligned} \tag{163}
$$

and

$$
\begin{aligned}
V_{\lambda}^{5} = &\frac{(m_A + m_B)}{2m_A m_B}\bar{\psi}_B(P_{\lambda} + {}_{\lambda}P)\psi_A h_2(t)g_5(t) \\
&+ \frac{(m_A - m_B)}{2m_A m_B}\bar{\psi}_B(P_{\lambda} - {}_{\lambda}P)\psi_A h_3(t)g_5(t).
\end{aligned} \tag{164}
$$

Thus we are left with the two independent vector current elements (158a) and (158b). That these two elements are independent and are just linear combinations of the 4-momenta is easily demonstrated. Explicitly we have that

$$\bar{\psi}_B(P_{\lambda})\psi_A = -\sqrt{m_B/m_A}\,\phi_B^*\partial_{\lambda}\phi_A, \tag{165}$$

$$\bar{\psi}_B({}_{\lambda}P)\psi_A = \sqrt{m_A/m_B}(\partial_{\lambda}\phi_B^*)\phi_A, \tag{166}$$

where ϕ_B^* and ϕ_A are KG fields.

We emphasize that for partially conserved currents, the DKP and KG matrix elements are inequivalent in the sense discussed earlier when considering minimal substitution and a partially conserved electric charge. Put simply, the difference arises because although the induced coupling t-dependence is essentially the same in the two formalisms, the different wave functions for the physical particle states have different t-dependences. The total t-dependence of the vector current matrix element is the combined dependence from both sources.

7. CONSERVED AND NONCONSERVED SPIN-1 CURRENTS

In this section we give a brief comparison of spin-1 currents in the Proca and DKP formalisms. The analysis is analogous to the discussion already given for spin-0 currents in Sec. 6.

For a free spin-1 particle of mass m the Proca equations are

$$\hat{U}_{\mu\nu} = \partial_{\mu}\hat{Q}_{\nu} - \partial_{\nu}\hat{Q}_{\mu}, \tag{167a}$$

$$\partial_{\mu}\hat{U}_{\mu\nu} = m^2\hat{Q}_{\nu} \tag{167b}$$

where $\hat{Q}_{\nu}$ is a 4-vector field $(\hat{\mathbf{Q}}, \hat{Q}_4)$ with each component being defined as a solution of the KG equation. $\hat{Q}_{\nu}$ also has imposed on it the subsidiary condition

$$\partial_{\nu}\hat{Q}_{\nu} = 0 \tag{168}$$

to remove timelike particles.

The DKP equations for spin-1 are

$$\partial_{\mu}\beta_{\mu}\psi = -m\psi, \tag{169a}$$

$$\partial_{\mu}\bar{\psi}\beta_{\mu} = m\bar{\psi}, \tag{169b}$$

with the β matrices being taken in the ten-dimensional representation. The connection between the Proca equations and the DKP equations is obtained by introducing the spin-1 projection operators of the R-algebra. In the ten-component DKP representation are four components $R_{\nu}\psi$ that transform as a 4-vector and six components $R_{\mu\nu}\psi$ which are the meson field strengths associated with the spin-1 field $R_{\nu}\psi$. Rewriting the DKP equation as coupled equations for these two sets of components, we have

$$\partial_{\lambda}(R_{\nu\lambda})\psi = -m(R_{\nu})\psi, \tag{170a}$$

$$\partial_{\nu}(R_{\lambda})\psi - \partial_{\lambda}(R_{\nu})\psi = m(R_{\nu\lambda})\psi. \tag{170b}$$

Defining

$$Q_{\nu} \equiv (R_{\nu})\psi\, m^{-1/2}, \tag{171}$$

the equations become

$$\partial_{\lambda}(R_{\nu\lambda})\psi = -m^{3/2}Q_{\nu}, \tag{172a}$$

$$\partial_{\nu}Q_{\lambda} - \partial_{\lambda}Q_{\nu} = m^{1/2}(R_{\nu\lambda})\psi. \tag{172b}$$

Thus

$$(R_{\nu\lambda})\psi = (1/m^{1/2})U_{\nu\lambda}, \tag{173}$$

where $U_{\nu\lambda}$ are the field strengths.

If the Proca field $\hat{Q}_\mu$ is taken to be complex the conserved charge current associated with the particle is

$$j_\mu^P = -i(\hat{Q}_\nu^* \hat{U}_{\mu\nu} - \hat{U}_{\mu\nu}^* \hat{Q}_\nu), \qquad \partial_\mu j_\mu^P = 0. \tag{174}$$

The corresponding DKP current is

$$j_\mu^{\mathrm{DKP}} = i\bar{\psi}\beta_\mu\psi, \quad \partial_\mu j_\mu^{\mathrm{DKP}} = 0, \tag{175}$$

which can be written in the V-algebra representation as

$$j_\mu^{\mathrm{DKP}} = i\sum_\lambda \bar{\psi}({}_\lambda V_{\lambda\nu} + {}_{\mu\lambda}V_\lambda)\psi. \tag{176}$$

Using the definition of the V-algebra elements in terms of the R-algebra elements, the DKP current can be written as

$$j_\mu^{\mathrm{DKP}} = i\sum_\lambda \bar{\psi}({}_\lambda R)(R_{\lambda\mu})\psi + i\sum_\lambda \bar{\psi}({}_{\mu\lambda}R)(R_\lambda)\psi. \tag{177}$$

From the definitions of $Q_\nu, U_{\mu\nu}$ the current is

$$j_\mu^{\mathrm{DKP}} = -i(Q_\nu{}^* U_{\mu\nu} - U_{\mu\nu}^* Q_\nu). \tag{178}$$

Hence for the free particle charge current the DKP expression for the current is identical to that obtained from the Proca equations.[17]

Considering now two particles of different masses, the Proca equations imply a transition current of the form

$$j_\mu^P = -i(\hat{Q}_\nu^{B\,*}\hat{U}_{\mu\nu}^A - \hat{U}_{\mu\nu}^{B\,*}\hat{Q}_\nu^A) \tag{179}$$

with divergence

$$\partial_\mu j_\mu^P = -i(m_A^2 - m_B^2)\hat{Q}_\nu^{B\,*}\hat{Q}_\nu^A. \tag{180}$$

The DKP transition current is

$$j_\mu^{\mathrm{DKP}} = i\bar{\psi}_B\beta_\mu\psi_A = i\sum_\lambda \bar{\psi}_B({}_\lambda R)(R_{\lambda\mu})\psi_A + i\sum_\lambda \bar{\psi}_B({}_{\mu\lambda}R)(R_\lambda)\psi_A \tag{181}$$

with divergence

$$\partial_\mu j_\mu^{\mathrm{DKP}} = i(m_B - m_A)\bar{\psi}_B\psi_A. \tag{182}$$

The DKP transition current can be cast in a Proca form as the free particle current was, yielding

$$j_\mu^{\mathrm{DKP}} = -i\left(\frac{m_B}{m_A}\right)^{1/2}(Q_\nu^B)^* U_{\mu\nu}^A + i\left(\frac{m_B}{m_A}\right)^{1/2}(U_{\mu\nu}^B)^* Q_\nu^A \tag{183}$$

Clearly the DKP transition current (183) is not the same as the Proca current (179). The DKP form of the current contains a term proportional to the Proca current plus an additional term.

ACKNOWLEDGMENT

Two of us (E.F. and M.M.N.) would like to thank the Aspen Center for Physics for its hospitality during the summer of 1972, when part of this work was done.

APPENDIX 1

The multiplication table for the P-algebra is

$$PP = P,$$

$$PP_\mu = P_\mu, \quad {}_\mu PP = {}_\mu P,$$

$$P({}_\mu P) = (P_\mu)P = P({}_\mu P_\nu) = ({}_\mu P_\nu)P = 0,$$

$$(P_\mu)(P_\nu) = ({}_\nu P)({}_\mu P) = 0,$$

$$({}_\mu P)(P_\nu) = {}_\mu P_\nu,$$

$$(P_\mu)({}_\nu P) = P\delta_{\mu\nu},$$

$$(P_\mu)({}_\nu P_\lambda) = P_\lambda\delta_{\mu\nu},$$

$$({}_\nu P_\lambda)({}_\mu P) = {}_\nu P\delta_{\lambda\mu},$$

$${}_\nu P({}_\mu P_\lambda) = ({}_\mu P_\lambda)P_\nu = 0,$$

$$({}_\mu P_\nu)({}_\sigma P_\lambda) = {}_\mu P_\lambda\delta_{\nu\sigma}.$$

The multiplication table for the V-algebra is

$$({}_\mu V_\nu)({}_\sigma V_\rho) = {}_\mu V_\rho\delta_{\nu\sigma},$$

$$({}_\mu V_\nu)({}_\sigma V_{\rho\tau}) = {}_\mu V_{\rho\tau}\delta_{\nu\sigma},$$

$$({}_\lambda V_\sigma)({}_{\rho\nu}V_\sigma) = ({}_\sigma V_{\rho\lambda})({}_\mu V_\nu) = 0,$$

$$({}_\mu V_\nu)({}_{\rho\lambda}V_{\sigma\tau}) = ({}_{\rho\lambda}V_{\sigma\tau})({}_\mu V_\nu) = 0,$$

$$({}_\mu V_{\nu\lambda})({}_{\rho\sigma}V_\tau) = {}_\mu V_\tau\Delta_{\nu\lambda\rho\sigma},$$

$$({}_{\mu\lambda}V_\gamma)({}_\sigma V_\rho) = {}_{\mu\lambda}V_\rho\delta_{\gamma\sigma}$$

$$({}_{\nu\lambda}V_\mu)({}_\tau V_{\rho\sigma}) = {}_{\nu\lambda}V_{\rho\sigma}\delta_{\mu\tau},$$

$$({}_\mu V_{\nu\lambda})({}_\tau V_{\rho\sigma}) = ({}_{\nu\lambda}V_\mu)({}_{\rho\sigma}V_\tau) = 0,$$

$$({}_\mu V_{\nu\lambda})({}_{\rho\sigma}V_{\tau\kappa}) = {}_\mu V_{\tau\kappa}\Delta_{\nu\lambda\rho\sigma},$$

$$({}_{\rho\sigma}V_{\tau\kappa})({}_{\nu\lambda}V_\mu) = {}_{\rho\sigma}V_\mu\Delta_{\tau\kappa\nu\lambda},$$

$$({}_{\rho\sigma}V_{\tau\kappa})({}_\mu V_{\nu\lambda}) = ({}_{\nu\lambda}V_\mu)({}_{\rho\sigma}V_{\tau\kappa}) = 0,$$

$$({}_{\rho\sigma}V_{\tau\kappa})({}_{\nu\lambda}V_{\mu\gamma}) = {}_{\rho\sigma}V_{\mu\gamma}\Delta_{\tau\kappa\nu\lambda}.$$

APPENDIX 2

In this appendix we give representations of the P- and V-subalgebras. Such representations can be obtained immediately from the multiplication rule for the independent basis elements of each of the algebras.

We had for the P-subalgebra that the multiplication rule for the linearly independent basis elements is

$$({}_aP_b)({}_cP_d) = ({}_aP_d)\delta(b,c).$$

We then have that a representation of the matrices is

$$({}_aP_b)_{mn} = \delta(m,a)\delta(n,b),$$

since

$$[({}_aP_b)({}_cP_d)]_{mn} = \delta(m,a)\delta(n,d)\delta(bc) = ({}_aP_d)_{mn}\delta(b,c).$$

With the DKP equation written in the Pauli metric, the representation of the P-algebra is

$$({}_iP_j)_{mn} = \delta(m,i+1)\delta(n,j+1), \quad i,j = 1,2,3,$$

$$({}_4P_4)_{mn} = \delta(m,1)\delta(n,1),$$

$$({}_5P_5)_{mn} = \delta(m,5)\delta(n,5),$$

$$({}_5P_i)_{mn} = \delta(m,5)\delta(n,i+1), \quad i = 1,2,3,$$

$$({}_4P_i)_{mn} = (-i)\delta(m,1)\delta(n,i+1), \quad i = 1,2,3,$$

$$({}_5P_4)_{mn} = (i)\delta(m,5)\delta(n,1),$$

$$({}_iP_5)_{mn} = ({}_5P_i)^*_{nm},$$

$$({}_iP_4)_{mn} = ({}_4P_i)^*_{nm},$$

$$({}_4P_5)_{mn} = ({}_5P_4)^*_{nm}.$$

For the V-algebra we had that the multiplication rule for the linearly independent basis elements is

$$({}_{ab}V_{cd})({}_{ef}V_{gh}) = {}_{ab}V_{gh}\Delta_{cdef},$$

where

$${}_{ab}V_{ca} = -{}_{ba}V_{cd} = {}_{ba}V_{dc}.$$

With $\{m, n = 1, \ldots, 10\}$ and $[cd]$ the ordered set $[12] \sim 1$, $[13] \sim 2$, $[14] \sim 3$, $[15] \sim 4$, $[23] \sim 5$, $[24] \sim 6$, $[25] \sim 7$, $[34] \sim 8$, $[35] \sim 9$, $[45] \sim 10$, a representation of the V-algebra is

$$({}_{ab}V_{cd})_{mn} = \delta'(m,[ab])\delta'(n,[dc]),$$

where the modified Kronecker δ-function is defined such that

$$\delta'(m,[ab]) = -\delta'(m,[ba])$$

and we choose

$$\delta'(m,[ab]) = \begin{cases} +1, & b > a \\ -1, & b < a \end{cases}.$$

Then, using the contraction

$$\sum_m \delta'(m,[ab])\delta'(m,[cd]) = \delta(a,c)\delta(b,d) - \delta(a,d)\delta(b,c)$$

we have that

$$\{({}_{ab}V_{cd})({}_{ef}V_{gh})\}_{mn} = \delta'(m,[ab])\delta'(n,[hg])\Delta_{cdef} = ({}_{ab}V_{gh})_{mn}\Delta_{cdef},$$

which verifies the representation.

*Supported in part by the Atomic Energy Commission (E.F. and M.M.N.) and the National Research Council of Canada (C.K.S.).

† Present address: Los Alamos Scientific Laboratory, Los Alamos, N.M. 87544.

[1]R. J. Duffin, Phys. Rev. **54,** 1114 (1938).
[2]N. Kemmer, Proc. R. Soc. A **173,** 91 (1939).
[3]G. Petiau, University of Paris, Thesis (1936). Published in Acad. R. Belg. Cl. Sci. Mem. Collect. 8 16 (2) (1936).
[4]I. Fujiwara, Prog. Theor. Phys. **10,** 589 (1953).
[5]Z. Tokuoka and H. Tanaka, Prog. Theor. Phys. 8, 599 (1952).
[6]Z. Tokuoka, Nucl. Phys. **78,** 681 (1966).
[7]T. Shimpuku, J. Math. Anal. Appl. **27,** 181 (1969); J. Math. Anal. Appl. **31,** 717 (1970).
[8]H. Umezawa, *Quantum Field Theory* (North-Holland, Amsterdam, 1956), p. 86.
[9]E. Fischbach, J. D. Louck, M. M. Nieto, and C. K. Scott, J. Math. Phys. (to be published).
[10]Harish-Chandra, Proc. R. Soc. A **186,** 502 (1946).
[11]E. Fischbach, F. Iachello, A. Lande, M. M. Nieto, and C. K. Scott, Phys. Rev. Lett. **26,** 1200 (1971).
[12]E. Fischbach, M. M. Nieto, H. Primakoff, C. K. Scott, and J. Smith, Phys. Rev. Lett. **27,** 1403 (1971).
[13]E. Fischbach, M. M. Nieto, H. Primakoff, and C. K. Scott, Phys. Rev. Lett. **29,** 1046 (1972).
[14]E. Fischbach, M. M. Nieto, and C. K. Scott, Prog. Theor. Phys. **48,** 574 (1972).
[15]E. Fischbach, M. M. Nieto, and C. K. Scott, Phys. Rev. D **6,** 726 (1972).
[16]E. Fischbach, M. M. Nieto, and C. K. Scott, Phys. Rev. D **7,** 207 (1973).
[17]A derivation along these lines was first given by S. N. Gupta, Proc. Cambridge Phil. Soc. 46, 649 (1950). Related earlier work was done by F. J. Belinfante, Physica 4, 849, 870, 887 (1939).

Diffusion, Einstein formula and mechanics

Gérard G. Emch

Department of Mathematics and of Physics, The University of Rochester, Rochester, New York 14627
(Received 11 April 1973)

A mathematically rigorous discussion of the diffusion equation and of its connection with the Einstein relation linking the diffusion coefficient and the velocity autocorrelation function is presented. Diffusion is then propounded as a typical case for which the logical consistency of a purely mechanistic theory of dissipative phenomena can be established.

INTRODUCTION

In this paper we analyze a diffusion process as a test case of a stationary dissipative system to be understood from the point of view of statistical mechanics. The literature on this subject abounds and one might justifiably be surprized that some serious problems of consistency still remain open. This seems however to be the case, although physicists admittedly accumulated a corpus of traditional wisdom in these matters. The purpose of this paper is precisely to show to which extent some of the semiheuristic arguments commonly found in the literature can be elevated to the status of mathematically meaningful statements. Specifically, we are interested in the mutual consistency of: (i) a diffusion equation valid for all positive times, (ii) an Einstein relation linking the diffusion coefficient to the time integral of the second derivative of the position covariance (i.e., to the time integral of the velocity autocorrelation function), and (iii) a purely mechanistic model in which the time evolution is conservative, i.e., can be described through a continuous, one-parameter unitary group.

We solve this problem as follows:

In Sec. 1 we review the phenomenological background for the diffusion equation we want to consider. This section is introduced for expository purposes only, and does not claim new results.

In Sec. 2 we analyze an argument originally due to Lebowitz, and show (Theorem 1) that it leads to some difficulties when applied to the situation described by our diffusion equation.

In Sec. 3 we present a Hilbert space solution of our diffusion equation. Specifically, we determine (Lemmas 1 and 2) the Hilbert space in which physics dictates that this equation be solved. We then show (Theorem 2) that our diffusion equation generates in this Hilbert space a contractive semigroup, the spectral properties of which we establish. Based on this solution, a description is then given (Theorem 3) of the analytic behavior of the position autocorrelation function (i.e., of the covariance of the stationary dissipative process defined by our diffusion equation). In particular, we show that the second derivative of this function diverges as $t^{-1/2}$ when t approaches + 0. We then derive rigorously an Einstein relation for the diffusion coefficient.

In Sec. 4 we show (Theorem 4) that the stationary dissipative system described by our diffusion equation admits purely mechanistic models and we establish the canonical spectral properties of the time-evolution for some particular "minimal" models which we define and show to exist. These minimal mechanical models are proven to be not only ergodic, but to be also mixing and even to have Lebesgue spectrum. We finally indicate (Theorem 5) some consequences of the just proven compatibility of the time-reversibility characteristic of mechanical systems, and of the irreversible behavior encountered in a stationary dissipative system.

The key results are thus isolated in the form of theorems, the proofs of which are given in an appendix.

1. PHENOMENOLOGICAL BACKGROUND

The aim of this section is to review the phenomenological basis of the theory.

We consider an ensemble of spherical particles, the movements of which are restricted to the positive real axis $\mathbb{R}^+$. Let $\rho(x,t)$ denote the density of these particles at (x,t) where x and t are, respectively, the position and time coordinates.

These particles are supposed to diffuse in some viscous medium at constant natural temperature $\beta(x,t) = \beta = (kT)^{-1}$, and to be submitted to an external potential $V(x,t) = mgx$. A more complicated space-dependance of the potential would only result in a somewhat more cumbersome notation, without affecting the main conclusions of this paper. We will therefore stick to this simple potential.

Let η, r, and v denote, respectively, the viscosity of the medium, the radius of the particles, and the instantaneous velocity of the particles.

For small values of r and v (specifically for $rv\rho\eta^{-1} \ll 1$), the Navier–Stokes equation leads to the Stokes law according to which the medium exercises on the moving particles a "frictional" force $F = -(6\pi r\eta)v$. We thus get the Langevin-type equation

$$m\dot{v} = -(6\pi r\eta)v - mg$$

from which we conclude that the particles reach a terminal velocity

$$-v_0 = -\mu mg, \qquad \text{where } \mu = (6\pi r\eta)^{-1}.$$

When the particles move at the terminal velocity $(-v_0)$ they produce a flow

$$\phi_S(x,t) = -v_0\rho(x,t) \equiv -\mu mg\rho(x,t).$$

On the other hand, the flux due to diffusion is given by the first Frick's law

$$\phi_D(x,t) = -D(\partial_x\rho)(x,t);$$

D will be referred to in the sequel as the *diffusion coefficient*. We thus obtain a net flow

$$\phi(x,t) = -v_0\rho(x,t) - D(\partial_x\rho)(x,t). \tag{1}$$

The next step is to invoke the conservation law for matter, which when expressed in differential form reads

$$(\partial_t\rho)(x,t) + (\partial_x\phi)(x,t) = 0. \tag{2}$$

Upon introducing (1) into (2) we get the *diffusion equation*

$$(\partial_t \rho)(x,t) = v_0(\partial_x \rho)(x,t) + D(\partial_x^2 \rho)(x,t). \tag{3}$$

The constraint that the particles are restricted to move on $\mathbb{R}^+$ means that an impenetrable wall is placed at $x = 0$, resulting in $\phi(0,t) = 0$ for all times t. We thus have to solve the partial differential equation (3) subject to the *boundary condition*

$$v_0\rho(0,t) + D(\partial_x \rho)(0,t) = 0 \quad \text{for all } t \in \mathbb{R}^+. \tag{4}$$

As is well known, Eq. (3) subjected to (4) and to

$$\int_0^\infty \rho(x,t)dx = 1 \quad \text{for all } t \in \mathbb{R}^+$$

admits a unique stationary solution

$$\rho_0(x,t) = (v_0/D)\exp[-(v_0/D)x] \tag{5}$$

which when compared to the canonical equilibrium density

$$\rho_0(x,t) = (\beta mg)\exp(-\beta mgx)$$

leads to the identification

$$D = \mu kT = kT/6\pi r\eta \tag{6}$$

known[1] as the *Nernst*[2]*–Einstein*[3] *formula.*

From the theory of Brownian motion,[4] viewed as a probabilistic approach to the diffusion equation, we only need to recall here that the average distance $\overline{\Delta}$ "travelled" by a diffusing particle during a time increment t is given by

$$\overline{\Delta}^2 = 2Dt. \tag{7}$$

Stricto Sensu, $\overline{\Delta}$ in this relation is to be interpreted as the dispersion at time t of a probability distribution concentrated on a point x at time $t = 0$. As is well known, relation (7) is verified experimentally over a very wide range of times.

A heuristic argument[1] seems to indicate that the velocity autocorrelation $\langle v(0)v(t)\rangle$ for such a particle is an even function of time, which essentially vanishes outside of a very small neighborhood of $t = 0$, and which is related to $\overline{\Delta}^2$ by

$$\overline{\Delta}^2 = t\int_{-\infty}^{+\infty} \langle v(0)v(t)\rangle dt. \tag{8}$$

Upon comparison of (7) and (8), one gets

$$D = \int_0^\infty \langle v(0)v(t)\rangle dt. \tag{9}$$

Although Einstein did not write this relation, it seems to underline part of his argument, and is thus referred to as the Einstein relation by some authors. For lack of a better term, we shall conform to this usage in this paper. We should also notice that (9) finds some supporting evidence, although still on a heuristic level, in the so-called linear response formalism.[1] From a formal point of view, however, one should realize that the derivation of (9) should be submitted to some further investigation. We shall come back to this problem later on in this paper.

2. THEORETICAL DIFFICULTIES

In spite of their incontestable successes in describing phenomena associated with diffusion, the considerations reviewed in the preceding section do leave open serious gaps as to their internal consistency, as well as to their consistency with mechanics.

As already pointed out by Einstein,[3] the velocity of a truly Brownian particle is a rather ill-defined concept. This is emphasized by the attempt to define $\bar{v}$ as the limit, as $t \to +0$, of $\overline{\Delta}/t$, which diverges to infinity as $t^{-1/2}$ if one uses (7): the "trajectory" of a truly Brownian particle would be a continuous, nowhere differentiable curve. The traditional wisdom, going back to Einstein, has it that mechanical considerations would prevent the diffusion equation to hold for "very small" times. Whereas we are willing to recognize the possibility that some departure from the diffusion equation might occur at "very small" times for some "real" physical systems, we shall prove, towards the end of this paper, that a diffusion equation, valid for *all* $t \in \mathbb{R}^+$ is *not* incompatible with the fundamental laws of mechanics.

An equally serious difficulty is brought to light by a careful analysis of an extremely interesting argument recently presented by Lebowitz.[5] We first want to go over Lebowitz' argument with a special emphasis of the assumptions under which it can be carried out. In doing so, we shall see that this argument can be extended sufficiently further than originally indicated so as to lead to the necessity of revising the otherwise quite reasonable assumptions underlying the original argument.

Let Σ be a (classical) mechanical system where $x_i(t)$, $v_i(t)$ denote respectively the position and the velocity of the ith particle at time t, and let us write $\langle\cdots\rangle$ for the canonical equilibrium average for some natural temperature $0 < \beta < \infty$. We say that Σ is *smooth* if the following conditions are satisfied.

(a) The position autocorrelation function ϕ defined by

$$\phi(T-t) = \langle x_i(t)x_i(T)\rangle$$

is continuous, and admits an absolutely continuous derivative;

(b) $\phi'(T-t) = -\langle v_i(t)x_i(T)\rangle = \langle x_i(t)v_i(T)\rangle$;

(c) $\phi''(T-t) = -\langle v_i(t)v_i(T)\rangle$;

(d) $\langle v_i(0)\rangle = 0 = \langle x_i(0)v_i(0)\rangle$;

(e) $\langle x_i(0)\rangle < +\infty$.

These conditions seem indeed to be reasonable if the Hamiltonian $H(x_1,\ldots,x_N,v_1,\ldots,v_N)$ of the system considered is of the form $H = T + V$, where

$$T = \Sigma_i(mv_i^2/2) \qquad V = V(x_1,\ldots,x_N)$$

and V comprises a "sufficiently smooth" interaction between the particles, *and* an external potential of the form

$$V_{\text{ext}} = \Sigma_i V_i(x_i)$$

with *either* $V_i(x_i) = 0$ for $0 \leq x_i \leq L$ and $V_i(x_i) = +\infty$ otherwise (i.e., Σ is a finite system enclosed within rigid walls); *or* $V_i(x_i) = \lambda_i x_i$ for $x_i \geq 0$ and $V_i(x_i) = +\infty$ otherwise (i.e., Σ is a semi-infinite system with a rigid wall at $x = 0$, and is submitted to an external homogeneous field).

In particular, we should point out that conditions (b) and (c) essentially express that

$$\frac{dx_i}{dt} = \frac{1}{m}\frac{\partial H}{\partial v_i} = v_i$$

and that d/dt and $\langle \ldots \rangle$ can be interchanged. As innocuous as these conditions might appear at this point we shall see presently that one of them at least has to be softened if our mechanical system is to show a diffusive behavior compatible with the Einstein formula. We shall show later on in this paper that such a modification is indeed required by the diffusion equation itself and that this modification is compatible with the fundamental laws of mechanics.

The following theorem extends an argument originally due to Lebowitz.[5]

Theorem 1: Let Σ be a smooth mechanical system.

(1) *If* Σ is *mixing* with respect to time, i.e., in particular,

$$\lim_{T\to\infty} |\langle a_i(0)b_i(T)\rangle - \langle a_i(0)\rangle\langle b_i(T)\rangle| = 0,$$

where $(a_i b_i)$ is any of the pairs (x_i, x_i), (x_i, v_i) and (v_i, v_i),

Then

$$D = \lim_{T\to\infty} \int_0^T \langle v_i(0)v_i(t)\rangle dt = 0.$$

(2) *If* Σ is only mixing with respect to time in configuration space, i.e., in particular,

$$\lim_{T\to\infty} |\langle x_i(0)x_i(T)\rangle - \langle x_i(0)\rangle\langle x_i(T)\rangle| = 0$$

Then either $\lim_{T\to\infty} \int_0^t \langle v_i(0)v_i(t)\rangle dt$ does not exist

or

$$D = \lim_{T\to\infty} \int_0^T \langle v_i(0)v_i(t)\rangle dt = 0.$$

(3) The conclusion of (2) can also be obtained, *without any mixing assumption, if* any *one* of the following conditions is satisfied

(i) $|\langle x_i(0)x_i(t)\rangle|$ is a bounded function of t,

(ii) $\langle x_i(0)^2\rangle < \infty$,

(iii) $\langle\{x_i(0) - \langle x_i(0)\rangle\}^2\rangle < \infty$,

(iv) Σ is of finite extension in configuration space.

The proof of this theorem can be found in the Appendix.

As already pointed out by Lebowitz, the third part of the theorem says in particular that a (smooth) mechanical system confined in a finite region of configuration space will never provide a model for diffusion in which one can recognize the occurence of a non vanishing diffusion coefficient D through the use of the Einstein relation (9).

Moreover, the formulation of the theorem which we gave shows that the same negative conclusion holds under more general situations. In particular, the condition that $0 \leq x_i \leq L < \infty$ can be released to the weaker condition that $\langle x_i(0)\rangle$ *and* $\langle x_i(0)^2\rangle$ be both finite, a situation which one might expect to be realized for a semi-infinite system in an homogeneous external field (see for instance Sec. 1 above).

This raises some serious questions of consistency. The first one is linked to the reliability of the Einstein relation. Specifically, one should ask whether the heuristic derivation of this formula can be straighted to a rigorous derivation, starting from the diffusion equation. We shall answer affirmatively this question in the next section. Anticipating on this result, we conclude that one at least of the "smoothness" conditions has to be released. We shall determine precisely which one in the course of the next section.

This suggests furthermore the question of whether the modification of the assumptions to be proposed in the next section will also invalidate the first and the second part of the theorem. We shall establish (also in the next section) that some mixing properties are indeed not only compatible with the diffusion equation, as one would like to expect from heuristic insight, but actually follow from this equation as rigorous consequences.

The last question suggested by the "no-go" theorem just given is that of the compatibility of the diffusion equation with the fundamental laws of mechanics. We shall also touch upon this question in Sec. 4.

3. SOLUTION OF THE DIFFUSION EQUATION

On the basis of the phenomenological considerations of Sec. 1, we have to solve the partial differential equation

$$\partial_t \rho = \Lambda_0 \rho \text{ with } \Lambda_0 = v_0 \partial_x + D\partial_x^2 \quad \text{and}$$

$$(x, t) \in \mathbb{R}^+ \times \mathbb{R}^+$$

subject to the boundary condition

$$v_0 \rho(0, t) + D(\partial_x \rho)(0, t) = 0 \quad \text{for all } t \in \mathbb{R}^+.$$

Our first step is to find a Hilbert space $\mathcal{K}_0$ on which this makes sense, both mathematically and physically.

In the sequel we denote by $\mathcal{C}_0(\mathbb{R}^+)$ the set of all bounded continuous functions $A : \mathbb{R}^+ \to \mathbb{C}$, equipped with the usual compositions laws with make $\mathcal{C}_0(\mathbb{R}^+)$ an Abelian C^*-algebra.

If ρ is a state on $\mathcal{C}_0(\mathbb{R}^+)$ we denote by $\langle \rho; A\rangle$ the expectation value of the observable $A \in \mathcal{C}_0(\mathbb{R}^+)$ in the state ρ. For instance, if

$$\rho_0(x) = (v_0/D) \exp[-(v_0/D)x],$$

$$\langle \rho_0; A\rangle = \int_0^\infty A(x)\rho_0(x)dx.$$

If ρ_1 and ρ_2 are two states on $\mathcal{C}_0(\mathbb{R}^+)$, we write $\rho_1 \dashv \rho_2$ whenever there exists a positive number λ such that $\langle \rho_1; A\rangle \leq \lambda\langle \rho_2; A\rangle$ for all positive $A \in \mathcal{C}_0(\mathbb{R}^+)$.

Finally, we mean by $\mathcal{L}^2(R^+, d\mu(x))$ the Hilbert space of all functions Ψ: $\mathbb{R}^+ \to \mathbb{C}$ which are square-integrable with respect to the measure μ on $\mathbb{R}^+$, i.e., such that

$$\|\Psi\|_\mu^2 = \int_0^\infty |\Psi(x)|^2 d\mu(x) < \infty.$$

As is usually done we drop the index μ when μ is the Lebesgue measure.

The following two well-known[6] lemmas provide a physical interpretation for our choice of the Hilbert space $\mathcal{K}_0$ in which we shall solve the diffusion equation stated in the beginning of this section.

Lemma 1: Let $\mathcal{K}_0 = \mathcal{L}^2(\mathbb{R}^+, d\mu(x))$ with $d\mu(x) = \exp(v_0 x/D)dx$, and $\mathcal{B}(\mathcal{K}_0)$ be the algebra of all bounded linear operators on $\mathcal{K}_0$. Then the mapping π: $\mathcal{C}_0(\mathbb{R}^+) \to \mathcal{B}(\mathcal{K}_0)$ defined by $(\pi(A)\Psi)(x) = A(x)\Psi(x)$ is a representation of the C^*-algebra $\mathcal{C}_0(\mathbb{R}^+)$ satisfying the following properties. There exists an element $\Phi_0 \in \mathcal{K}_0$ such that

(i) $(\Phi_0, \pi(A)\Phi_0)_\mu = \langle \rho_0; A\rangle$ for all $A \in \mathcal{C}_0(\mathbb{R}^+)$,

(ii) $\{\pi(A)\Phi_0 \mid A \in \mathcal{C}_0(\mathbb{R}^+)\}$ is dense in $\mathcal{H}_0$.

Specifically, $\Phi_0(x) = (v_0/D)^{1/2} \exp(-v_0 x/D)$. Up to unitary equivalence π is the only representation of $\mathcal{C}_0(\mathbb{R}^+)$ satisfying properties (i) and (ii) above.

Lemma 2: To any state ρ on $\mathcal{C}_0(\mathbb{R}^+)$ such that $\rho \rightharpoonup \rho_0$ corresponds a unique positive element $B \in \{\pi(\mathcal{C}_0(\mathbb{R}^+))\}''$ such that $\langle \rho; A\rangle = (\Phi_0, \pi(A)B\Phi_0)_\mu$ for all $A \in \mathcal{C}_0(\mathbb{R}^+)$. Conversely, this formula defines a state ρ on $\mathcal{C}_0(\mathbb{R}^+)$ with $\rho \rightharpoonup \rho_0$ whenever B is a positive element in $\{\pi(\mathcal{C}_0(\mathbb{R}^+))\}''$ such that $(\Phi_0, B\Phi_0)_\mu = 1$.

For the sake of notational simplicity, and when no confusion is likely to occur, we shall write A for $\pi(A)$. For the same reason, we shall simply write $\mathcal{C}_0(\mathbb{R}^+)''$ for the (maximal Abelian) von Neumann algebra $\{\pi(\mathcal{C}_0(\mathbb{R}^+))\}''$ generated in $\mathcal{B}(\mathcal{H}_0)$ by $\pi(\mathcal{C}_0(\mathbb{R}^+))$, and we will identify it with the algebra $\mathcal{L}^\infty(\mathbb{R}^+, dx)$ of all essentially bounded functions on $\mathbb{R}^+$ with respect to Lebesgue measure. Similarly, we shall simply denote by ρ_0 the unique normal extension $\tilde{\rho}_0$ of ρ_0 to $\mathcal{C}_0(\mathbb{R}^+)''$, and thus write for instance the conclusion of Lemma 2 as $\langle \rho; A\rangle = \langle \rho_0; AB\rangle$.

We now assert the integrability of our diffusion equation within the space $\mathcal{H}_0 = \mathcal{L}^2(\mathbb{R}^+, d\mu(x))$ of lemmas 1 and 2 above, and we give some basic properties of its solutions:

Theorem 2: Let $\mathcal{D}(\Lambda_0)$ be the dense linear manifold in $\mathcal{H}_0$ constituted by all $\Psi \in \mathcal{H}_0$ such that (i) Ψ has absolutely continuous first derivative, and (ii) $v_0 d_x\Psi + Dd_x^2\Psi \in \mathcal{H}_0$, (iii) $v_0\Psi(0) + D(d_x\Psi)(0) = 0$. Let Λ_0 be the operator defined on $\mathcal{D}(\Lambda_0)$ by $\Lambda_0 = v_0 d_x + Dd_x^2$. With $\Phi_0(x) = (v_0/D)^{1/2} \exp(-v_0 x/D)$, let $\mathcal{H}_0^{(0)}$ (resp. $\mathcal{H}_0^{\perp}$) denote the one-dimensional subspace of $\mathcal{H}_0$ generated by Φ_0 (resp. the orthocomplement of $\mathcal{H}_0{}^{(0)}$ in $\mathcal{H}_0$). *Then*

(a) Λ_0 admits only one eigenvalue, namely zero, and the corresponding eigensubspace is $\mathcal{H}_0{}^{(0)}$.

(b) The restriction $\Lambda_0^{\perp}$ of Λ_0 to $\mathcal{D}(\Lambda_0) \cap \mathcal{H}_0^{\perp}$ is unitarily equivalent to the self-adjoint operator Λ defined in $\mathcal{H} = \mathcal{L}^2(\mathbb{R}^+, dx)$ by $\Lambda = -(v_0^2/4D)\, I - Dx^2$ on the natural domain $\mathcal{D}(\Lambda) = \{\Psi \in \mathcal{H} \mid \int_0^\infty x^4 |\Psi(x)|^2\, dx < \infty\}$.

(c) Λ_0 is self-adjoint. Its spectrum $Sp(\Lambda_0)$ is the union of a discrete part $Sp_d(\Lambda_0) = \{0\}$ and a part $Sp_{ac}(\Lambda_0)$ which is absolutely continuous with respect to Lebesgue measure, and extends from $-\infty$ to $-(v_0^2/4D)$. $Sp(\Lambda_0)$ is simple.

(d) Λ_0 is the generator of a contracting semigroup $\{S_0(t) \mid t \in \mathbb{R}^+\}$ of self-adjoint operators on $\mathcal{H}_0$. In particular, if $S_0{}^{(0)}(t)$ and $S_0^{\perp}(t)$ denote, respectively, the restrictions of $S(t)$ to the invariant subspaces $\mathcal{H}_0{}^{(0)}$ and $\mathcal{H}_0^{\perp}$, we have $S_0{}^{(0)}(t) = I$ and $\|S_0^{\perp}(t)\| \leq \exp(-v_0^2 t/4D)$.

(e) For every $\rho(t)$ with $\rho(0) \rightharpoonup \rho_0$ satisfying the diffusion equation, there exists a unique positive element $B \in \mathcal{C}_0(\mathbb{R}^+)''$ such that $\langle \rho(t); A\rangle = (\Phi_0, AS(t)B\Phi_0)_\mu$.

The proof of this theorem is sketched in the Appendix.

Upon using the generalized Fourier transform[7] which establishes the unitary equivalence between Λ_0 acting in $\mathcal{H}_0$ and the operator $0 \oplus \Lambda$ acting in $\mathbb{C} \oplus \mathcal{H}$, we can in particular explicitly compute, for every t in $\mathbb{R}^+$, the numerical value of $\langle \rho(t); A\rangle$ for every A in $\mathcal{C}_0(\mathbb{R}^+)$ whenever the initial state of Σ satisfies the condition $\rho \rightharpoonup \rho_0$.

Moreover, as a direct consequence of part (d) of the above theorem, we see immediately that

$$\langle \rho(t); A\rangle = \langle \rho_0; A\rangle + (\Phi_0, A(I - E_0)S_0^{\perp}(t)(I - E_0)B\Phi_0)_\mu,$$

where E_0 is the projector from $\mathcal{H}_0$ onto $\mathcal{H}_0{}^{(0)}$. In particular,

$$|\langle \rho(t); A\rangle - \langle \rho_0, A\rangle| \leq \|A\Phi_0\|_\mu \cdot \|B\Phi_0\|_\mu \exp(-v_0^2 t/4D)$$

so that $\rho(t)$ tends exponentially (in the w^*-topology) to the equilibrium distribution ρ_0 as t tends to infinity. This illustrates the dissipative character of our diffusion equation.

In a general system the time evolution would be given by $S(t) = \exp[-i(\Omega + i\Lambda)t]$ with Ω and Λ self-adjoint. If $\Lambda = 0$, the time evolution would be unitary and the system would be a conservative, mechanical system. Here the exact opposite occurs: $\Omega = 0$ and $\Lambda = \Lambda_0 \neq 0$. It should be realized that the extreme dissipativeness expressed by the self-adjointness and negative definiteness of the generator $(v_0 d_x + Dd_x^2)$ of the time evolution is intimately linked to the measure $d\mu(x) = \exp[v_0 x/D]$ imposed on us by our problem.

It will be our task in the next section to show that this extreme dissipative behavior is nevertheless strictly compatible with a global mechanical evolution.

We presently want to exploit the spectral properties of the diffusion equation which are established in Theorem 2 and compare their consequences to the assumptions and results of Secs. 1 and 2.

Lemma 3: *Let* $\{S_0(t) \mid t \in \mathbb{R}^+\}$ be the semigroup generated in $\mathcal{H}_0$ by our diffusion equation, and Λ_0 be its self-adjoint generator. *Then* for every $t > 0$ and every integer $n > 0$

(a) $S_0(t)\mathcal{H}_0 \subset \mathcal{D}(\Lambda_0)$,

(b) for every Φ and Ψ in $\mathcal{H}_0$, $(\Phi, S_0(t)\Psi)_\mu$ is continuously differentiable to all orders and there exist $A_n < \infty$ such that

$$|d_t^n(\Phi, S_0(t)\Psi)_\mu| \leq A_n t^{-n}.$$

The proof of this lemma is given in the Appendix.

Theorem 3: *Let* X be the operator defined on the natural domain $\mathcal{D}(X) \subset \mathcal{H}_0$ by $(X\Psi)(x) = x\Psi(x)$. *Then*

(a) $\phi(t) \equiv (\Phi_0, XS_0(t)X\Phi_0)_\mu$ is a continuous bounded function of $t \in \mathbb{R}^+$;

(b) $\lim_{t \to +0} \phi(t) = 2(D/v_0)^2 = \langle \rho_0; X^2\rangle$,

$\lim_{t \to +\infty} \phi(t) = (D/v_0)^2 = (\Phi_0, X\Phi_0)_\mu^2 = \langle \rho_0; X\rangle^2$,

(c) for any integer $n > 0$ the nth derivative $\phi^{(n)}(t)$ of $\phi(t)$ exists and is bounded on any closed semibounded interval of $\mathbb{R}^+$ excluding $t = 0$, and

$$\lim_{t \to +\infty} \phi^{(n)}(t) = 0;$$

(d) $\lim_{t \to +0} \phi'(t) = -D$,

$$\lim_{t \to +0} |\phi''(t) - v_0(D/\pi)^{1/2} t^{-1/2}| = v_0^2.$$

The proof of this theorem is established in the Appendix.

We now turn to the physical consequences of our diffusion equation. In particular, we want to use the firm mathematical basis provided by the above results to discuss the question of whether or not some of the semi-heuristic assertions of Secs. 1 and 2 are compatible with this equation.

We already noticed the fact that $\rho(0) \dashv \rho_0$ implies that $\rho(t)$ approaches exponentially ρ_0 as t tends to infinity. By Lemma 2, this result can be equivalently stated as saying that the following exponential mixing property holds; namely, that for every configurational observables A and B in $\mathcal{C}_0(\mathbb{R}^+)$ [or even in $\mathcal{C}_0(\mathbb{R}^+)''$],

$$|\langle\rho_0; AB(t)\rangle - \langle\rho_0; A\rangle\langle\rho_0; B\rangle| \leq \|A\|\cdot\|B\|\cdot \exp(-v_0^2 t/4D),$$

where $\langle\rho_0; AB(t)\rangle = (\Phi_0, AS_0(t)B\Phi_0)_\mu$.

Moreover, we see from Theorem 3 that a similar exponential mixing property still holds for the position autocorrelation function obtained by substituting in the above expression the unbounded observable X to the bounded observables A and B. Consequently, *if* we want to adhere to a description compatible with the diffusion equation, and *if* we want to interpret (as suggested in Sec. 2) $\phi(t)$, $-\phi'(t)$, and $-\phi''(t)$ as, respectively, $\langle x(0)x(t)\rangle$, $\langle v(0)x(t)\rangle$, and $\langle v(0)v(t)\rangle$, *then* Theorem 1 requires that either the Einstein formula or some smoothness assumption must default. Upon looking at the results of Theorem 3, we see immediately which of the assumptions of Theorem 1 is in conflict with the diffusion equation: $\langle v(0)v(t)\rangle$ is compatible with the diffusion equation if and only if it diverges as $t^{-1/2}$ when t approaches 0. This singular behavior of $\langle v(0)v(t)\rangle$ also manifests itself, though in a milder form, in the behavior of $\langle v(0)x(t)\rangle$. Whereas this function is absolutely continuous on any interval $[\epsilon, \infty)$ with $\epsilon > 0$, and its limit exists as t tends to $+0$, that limit however is equal to D, not to zero. This fact is of primary importance in relation with the Einstein formula. Indeed, upon using the absolute continuity of $\langle v(0)x(t)\rangle$ and the fact that this function tends to zero as t tends to infinity, *we derive the Einstein relation as a consequence of the diffusion equation* (where v_0 and D are non-zero). Specifically, it follows from Theorem 3 that

$$\lim_{\epsilon\to+0}\lim_{T\to+\infty}\int_\epsilon^T \phi''(t)dt = \lim_{T\to+\infty}\phi'(T) - \lim_{\epsilon\to+0}\phi'(\epsilon) = D, \quad (10)$$

i.e., if we substitute in this mathematically rigorous result the above mentioned interpretation of $\phi''(t)$ for $t > 0$, then

$$D = -\lim_{\epsilon\to+0}\lim_{T\to+\infty}\int_\epsilon^T \langle v(0)v(t)\rangle dt. \quad (11)$$

We might notice at this point that the Einstein formula derived above is independent from the numerical value of v_0, provided that v_0 be different from zero. This makes trivial a subsequent limiting procedure in which v_0 would approach zero. This evidently confirms the physical expectation that, within the limits of the theory behind the diffusion equation, the value of the diffusion constant, as computed from the Einstein relation, is independent of the strength of the external field. It should nevertheless be pointed out that the limit in which v_0 tends to zero should not be taken from the onset in a (semi-) infinite system, since the essential singularity of $\langle v(0)v(t)\rangle$ would then become catastrophic: We would indeed have then that $\langle v(0)v(t)\rangle$ is identically zero for all $t > 0$. The same conclusion is reached if one approaches the situation $v_0 = 0$ for the semi-infinite system by letting L tend to infinity in the following diffusive system:

$$\partial_t\rho = \Lambda\rho \quad \text{with } \Lambda = D\partial_x^2 \quad \text{and} \quad (x,t) \in [0,L]\times\mathbb{R}^+$$

subject to the boundary conditions

$$(\partial_x\rho)(0,t) = 0 = (\partial_x\rho)(L,t) \quad \text{for all} \quad t \in \mathbb{R}^+.$$

The reader will check that in this case (i) the diffusion operator Λ has simple, discrete spectrum $Sp(\Lambda) = \{-(n\pi/L)^2 D \,|\, n \in Z^+\}$, (ii) the stationary distribution becomes $\rho_0(x) = L^{-1}$, and (iii) the position autocorrelation function is

$$\langle x(0)x(t)\rangle = L^2\Big((1/4) + (8/\pi^4)\sum_{m=0}^{\infty}(2m+1)^{-4} \times \exp[-(2m+1)^2(\pi^2 D)(t/L^2)]\Big)$$

from which we conclude again, *mutatis mutandis*, to the validity of Theorem 3, and in particular to the validity of relation (10), and to the occurrence of a divergence of $\phi''(t)$ as t approaches zero, whereas $\phi''(t)$ approaches zero exponentially fast as t tends to infinity. [We notice, incidentally, that the diffusive cases just discussed appear actually to be only two particular examples of stationary dissipative systems for which the relation (10) holds in general.] This behavior of $\phi''(t)$ is reflected in the fact that in the limit where L tends to infinity $\langle v(0)v(t)\rangle$ is identically zero for all $t > 0$, and diverges to infinity at $t = 0$. This confirms the result obtained above when this physical situation is approached as the limit of a semi-infinite system where v_0 tends to zero. This fact, thus established from two different approaches, provides a rigorous basis for the *ad hoc* assumption that $\langle v(0)v(t)\rangle \sim \delta(t)$ encountered in some[1] heuristic derivations of the Einstein relation when $v_0 = 0$.

We now return to the case of a semi-infinite system where $v_0 > 0$, which is the proper domain of the investigation carried out in this paper. We can summarize the results obtained in this section by the following conclusions.

Starting from the diffusion equation in a semi-infinite medium and in the presence of an homogeneous external field, we proved that the position autocorrelation function $\phi(t)$ [$= \langle x(0)x(t)\rangle = \langle\rho_0, XX(t)\rangle = (\Phi_0, XS_0(t)X\Phi_0)_\mu$] and all its derivatives behave indeed as *assumed* in Theorem 1 when t tends to infinity. In addition, we proved that $\phi''(t)$ [$= -\langle v(0)v(t)\rangle$] exhibits an essential singularity at $t = 0$, which violates the assumption of Theorem 1, and is responsible, indeed inseparable, from the nonvanishing of the diffusion coefficient D.

We should finally remark that we only addressed ourselves in this section to the problem of the derivation and interpretation of an Einstein relation compatible with the diffusion equation, namely formula (10) above where $\phi(t) = \langle\rho_0, XX(t)\rangle$. This relation is now established as a direct consequence of the diffusion equation. Formula (11) on the other hand is of a less fundamental character. *Stricto sensu* it should only be considered as a rewriting of (10) in a suggestive notation. Moreover, the compatibility of the diffusion equation with the fundamental laws of mechanics will only be examined in the next section. We can nevertheless affirm at this point that a strict adherence to the diffusion equation imposes that the respective roles of the upper ($T \to +\infty$) and lower ($\epsilon \to +0$) limits appearing in the Einstein formula (10) run at countercurrent to some recent attempts[5,8-10] to check the occurrence or the absence of self-diffusion through an Einstein formula of the type of (11) where the one-particle velocity autocorrelation function

$\langle v_i(0)v_i(t)\rangle$ plays the role of $\phi''(t)$. It might therefore be hoped that the present investigation provide a supplementary tool to elucidate the interpretation and true relevance of mechanistic models for diffusive processes.

4. DIFFUSION EQUATION AND MECHANICS

In this section we first define precisely what we mean in general by a mechanical system, by a stationary dissipative system and by the assertion that a dissipative system can be embedded as a subsystem of a mechanical system. We then show that the stationary dissipative system defined by the diffusion equation studied in the preceding section can be canonically embedded in a minimal mechanical system, and we study the spectral properties of the evolution for this mechanical system.

A *mechanical system* is defined here as a triple $(\mathfrak{A}, \{\alpha_\phi(t) \mid t \in \mathbb{R}\}, \phi)$ formed by a C^*-algebra $\mathfrak{A}$, a state ϕ on $\mathfrak{A}$, and a weakly-continuous, one-parameter group $\{U_\phi(t) \mid t \in \mathbb{R}\}$ of unitary operators acting on the Hilbert space $\mathcal{H}_\phi$ canonically associated to ϕ by the GNS construction, and such that

(i) $U_\phi(t)\Phi = \Phi$ for all $t \in \mathbb{R}$,

(ii) $U_\phi(t)\pi_\phi(A)U_\phi(-t)$ belongs to $\pi_\phi(\mathfrak{A})''$ for all $A \in \mathfrak{A}$ and all $t \in \mathbb{R}$.

We denote by $\{\alpha_\phi(t) \mid t \in \mathbb{R}\}$ the group automorphisms of $\pi_\phi(\mathfrak{A})''$ generated by $\{U_\phi(t) \mid t \in \mathbb{R}\}$.

For the intelligence of the above definition we recall[6] that Lemma 1 can be extended to an arbitrary C^*-algebra $\mathfrak{A}$ as follows.

Lemma 4: (GNS construction): To every state ϕ on a C^*-algebra $\mathfrak{A}$ corresponds a triple $(\mathcal{H}_\phi, \pi_\phi(\mathfrak{A}), \Phi)$ consisting of a Hilbert space $\mathcal{H}_\phi$, a representation π_ϕ of $\mathfrak{A}$ into the algebra $\mathfrak{B}(\mathcal{H}_\phi)$ of all bounded linear operators on $\mathcal{H}_\phi$, and a vector $\Phi \in \mathcal{H}_\phi$ such that (i) $(\Phi, \pi_\phi(A)\Phi) = \langle \phi; A\rangle$ for all $A \in \mathfrak{A}$, and (ii) $\{\pi_\phi(A)\Phi \mid A \in \mathfrak{A}\}$ is dense in $\mathcal{H}_\phi$. Moreover, $(\mathcal{H}_\phi, \pi_\phi(\mathfrak{A}), \Phi)$ is uniquely determined, up to unitary equivalence, by the above conditions.

As in Sec. 3 when no confusion is likely to occur we write simply A for $\pi_\phi(A)$, $\mathfrak{A}''$ for $\pi_\phi(\mathfrak{A})''$, $\mathcal{H}$ for $\mathcal{H}_\phi$, $U(t)$ for $U_\phi(t)$, and $\alpha(t)$ for $\alpha_\phi(t)$.

Our definition of a mechanical system supposes thus the existence of a time invariant state ϕ, and that the time evolution, unitarily implemented in the representation of $\mathfrak{A}$ canonically associated to ϕ, transforms the observables in $\mathfrak{A}''$ amongst themselves, while preserving their algebraic relations. This definition covers the case of a classical flow.[11] The technical assumption that $\alpha(t)$ be an automorphism of $\mathfrak{A}''$ rather than $\mathfrak{A}$ is introduced for reasons on which we will not say more here than that is forced upon us by the consideration of various models[9,12,13] in statistical mechanics.

A *stationary dissipative system* is defined here as a triple $(\mathfrak{A}_0, \{S_0(t) \mid t \in \mathbb{R}^+\}, \phi_0)$ formed by a C^*-algebra $\mathfrak{A}_0$, a state ϕ_0 on $\mathfrak{A}_0$, and a weakly-continuous, one-parameter semigroup $\{S_0(t) \mid t \in \mathbb{R}^+\}$ of contractions acting on the Hilbert space $\mathcal{H}_0$ canonically associated to ϕ_0 by the GNS construction, and such that

(i) $S_0(t)\Phi_0 = \Phi_0$ for all $t \in \mathbb{R}^+$,

(ii) $\langle \phi_0, AB(t)\rangle \equiv (\Phi_0, AS_0(t)B\Phi_0)$.

Theorem 2 precisely gives an example of a stationary dissipative system, namely, $(\mathcal{C}_0(\mathbb{R}^+), \{S_0(t) \mid t \in \mathbb{R}^+\}, \phi_0)$, where

$$\langle \phi_0; A\rangle = \int_0^\infty dx A(x)\rho_0(x)$$

with $\rho_0(x) = (v_0/D)\exp(-v_0x/D)$

and where $\{S_0(t) \mid t \in \mathbb{R}^+\}$ is the semigroup generated by the diffusion equation

$$\partial_t\rho(x,t) = (v_0\partial_x + D\partial_x^2)\rho(x,t)$$

with boundary condition

$$v_0\rho(0,t) + D(\partial_x\rho)(0,t) = 0 \quad \text{for all } t \in \mathbb{R}^+.$$

A mechanical system $(\mathfrak{A}, \{\alpha(t) \mid t \in \mathbb{R}\}, \phi)$ and a stationary dissipative system $(\mathfrak{A}_0, \{S_0(t) \mid t \in \mathbb{R}^+\}, \phi_0)$ are said to be *compatible* if

(i) $\mathfrak{A}_0$ is a sub C^*-algebra of $\mathfrak{A}$,

(ii) $\langle \phi; A\alpha(t)[B]\rangle = \langle \phi_0; AB(t)\rangle$ for all $t \in \mathbb{R}^+$ and all A and B in $\mathfrak{A}_0$.

We also refer to this situation by saying that $(\mathfrak{A}, \{\alpha(t) \mid t \in \mathbb{R}\}, \phi)$ is a mechanical model for $(\mathfrak{A}_0, \{S_0(t) \mid t \in \mathbb{R}^+\}, \phi_0)$, or that the time-evolution of the mechanical system $(\mathfrak{A}, \{\alpha(t) \mid t \in \mathbb{R}\}, \phi)$ when restricted to $\mathfrak{A}_0 \subset \mathfrak{A}$ generates a stationary dissipative system.

Suppose now that $(\mathfrak{A}, \{\alpha(t) \mid t \in \mathbb{R}\}, \phi)$ and $(\mathfrak{A}_0, \{S_0(t) \mid t \in \mathbb{R}^+\}, \phi_0)$ are compatible. Because of Lemma 4, we can identify Φ_0 with Φ, and $\mathcal{H}_0$ with the subspace obtained as the closure of $\mathfrak{A}_0\Phi_0$ in $\mathcal{H}$. We say that $(\mathfrak{A}, \{\alpha(t) \mid t \in \mathbb{R}\}, \phi)$ is *minimal* with respect to $(\mathfrak{A}_0, \{S_0(t) \mid t \in \mathbb{R}^+\}, \phi_0)$ if $\{\alpha(t)[A]\Phi_0 \mid t \in \mathbb{R}, A \in \mathfrak{A}_0\}$ is dense in $\mathcal{H}$.

We now state the two results of this section, the proof of which are given in the Appendix.

Theorem 4: The stationary dissipative system $(\mathcal{C}_0(\mathbb{R}^+), \{S_0(t) \mid t \in \mathbb{R}^+\}, \phi_0)$ generated by the diffusion equation $\partial_t\rho(x,t) = (v_0\partial_x + D\partial_x^2)\rho(x,t)$ with boundary condition $v_0\rho(0,t) + D(\partial_x\rho)(0,t) = 0$ for all $t \in \mathbb{R}^+$, admits a minimal mechanical model. The spectrum $Sp(H)$ of the generator H of the unitary group $\{U(t) \mid t \in \mathbb{R}\}$ describing the time-evolution in any minimal mechanical model of $(\mathcal{C}_0(\mathbb{R}^+), \{S_0(t) \mid t \in \mathbb{R}^+\}, \phi_0)$ is uniquely determined by the stationary dissipative system itself. It is the union of a discrete part $Sp_d(H) = \{0\}$ which is nondegenerate, and of $Sp_{ac}(H) = \mathbb{R}$ which is absolutely continuous with respect to Lebesgue measure and is countably infinitely degenerated.

This theorem thus asserts the existence of mechanical models for our diffusion equation in the precise sense defined above. Moreover, it asserts the existence minimal mechanical models. Furthermore, these minimal mechanical models are not only ergodic but they are also mixing and even have Lebesgue spectrum.

Theorem 5: In any mechanical model of our diffusion equation

$$\langle \phi; A\alpha_t[B]\rangle = \langle \phi; A\alpha_{-t}[B]\rangle = \langle \phi_0; AB(|t|)\rangle$$

for all $t \in \mathbb{R}$ and all A and B in $\mathfrak{A}_0$. Moreover, the only vectors in $\mathcal{H}_0$ on which the generator H of the mechanistic time-evolution is defined are of the form $\lambda\Phi_0$ with $\lambda \in \mathbb{C}$.

The first part of this theorem illustrates the compatibility of the time-reversibility characteristic of mechanical systems, and of the irreversible behavior encountered in dissipative systems. The second part of the theorem indicates the source of some of the pecularities, such as divergent diagrams, often appearing in attempts[14] to derive irreversible equations from mechanistic assumptions. Indeed every term of a formal power expansion of $(\Phi_0, A_1(t_1), \ldots, A_n(t_n)\Phi_0)$ with $A_1, .., A_n \in \mathfrak{A}_0$ would contain terms of the form $(\Phi_0, A_1 H \ldots H A_n \Phi_0)$, where $HA_n\Phi_0$ is not defined except for the trivial case when $A\Phi_0 = a\Phi_0$. Moreover, the fact that $\mathfrak{D}(H) \cap \mathcal{H}_0 = \{\lambda\Phi_0 | \lambda \in \mathbb{C}\}$ implies that $\alpha(t)[X]\Phi_0 = U(t)XU(-t)\Phi_0 = U(t)X\Phi_0$ does not admit a derivative at $t = 0$, so that a velocity operator $V(0)$ can actually not be defined on Φ_0. This is evidently linked to the divergence of $\Phi''(t)$ as t approaches zero encountered in the preceding section, and which we interpreted as violating one of the questionable smoothness assumption often made about mechanical systems.

5. CONCLUSIONS

We gave (in Sec. 4) a generalized definition of mechanical systems, and proved that a mechanical system satisfying this definition can be canonically generated from the stationary dissipative system corresponding to a diffusion equation valid for all positive times.

We proved that the time-evolution for the mechanical system so obtained has Lebesgue spectrum.

We showed how and why this mechanical system cannot, however, satisfy some of the smoothness assumptions often made on the more restrictive mechanical systems described in Sec. 2. These restrictive assumptions are not used in the subsequent sections, thus allowing us to indicate precisely which of the smoothness assumptions should be modified, and to give explicitly the weakened version of these assumptions which is satisfied.

In the framework imposed by a strict adherence to the diffusion equation, we gave a mathematically sound derivation, and thus a reliable interpretation, of the Einstein relation [our formula (10)].

On the firm mathematical basis provided by the analysis carried out in this paper we can further speculate that the time-scale renormalization of a smooth mechanical system often[15] invoked to derive a diffusion equation will result in a dynamical system which will still satisfy the general definition of a mechanical system proposed in Sec. 4 and the weakened smoothness properties of the time-evolution indicated in Sec. 3, and which will admit a reduced description by a diffusion equation generating a stationary dissipative system in the sense of Sec. 4.

ACKNOWLEDGMENTS

At various stages in the elaboration of this paper, the author benefited from helpful conversations with Mr. T. Ali, Dr. J. Keilson, Dr. O. Lanford, Dr. R. Lavine, and Dr. J. L. Lebowitz.

APPENDIX

Proof of Theorem 1: We first notice that the smoothness conditions imposed on Σ imply that

$$\int_0^T \langle v_i(0)v_i(t)\rangle dt = -\int_0^T \phi''(t)dt = -\phi'(T) + \phi'(0)$$
$$= \langle v_i(0)x_i(T)\rangle - \langle v_i(0)x_i(0)\rangle = \langle v_i(0)x_i(T)\rangle.$$

It is therefore sufficient for the purpose of establishing the theorem to study the limit as T tends to infinity, of $\langle v_i(0)x_i(T)\rangle$. If Σ is mixing, we have in particular

$$\lim_{T\to\infty} \langle v_i(0)x_i(T)\rangle = \langle v_i(0)\rangle\langle x_i(0)\rangle = 0$$

thus proving the first part of the theorem. To prove the second part of the theorem, we first notice that

$$\langle v_i(0)x_i(t)\rangle = -\phi'(t) = -\frac{d}{dt}\langle x_i(0)x_i(t)\rangle.$$

If now Σ is mixing in configuration space, we have, since on the other hand $\langle x_i(0)\rangle = \langle x_i(t)\rangle$ is finite, that

$$\lim_{T\to\infty} \phi(T) = \lim_{T\to\infty} \langle x_i(0)x_i(T)\rangle = \langle x_i(0)\rangle^2$$

which is finite. Consequently, either the limit, as T tends to infinity, of $\phi'(T)$ does not exist, or it exists and is zero. Upon using the remark opening this proof, we see that the second part of the theorem is now proven. To prove the third part of the theorem, we notice that if (iv) is satisfied, then so is trivially (iii). Moreover, since $\langle x_i(0)\rangle$ is finite, (iii) and (ii) are equivalent. Finally, by Schwartz inequality, and the fact that $\langle x_i(0)\rangle = \langle x_i(t)\rangle$ is finite, (ii) implies (i). Now, (i) says that $|\phi(t)|$ is bounded. Consequently, $\phi'(T)$ either approaches zero as T tends to infinity, or it does not approach any limiting value. Joined to the remark opening the proof of the first part of the theorem, this completes the proof of the theorem. ■

Proof of Theorem 2: The unitary transformation U: $\mathcal{H} = \mathcal{L}^2(\mathbb{R}^+, dx) \to \mathcal{H}_0 = \mathcal{L}^2(\mathbb{R}^+, d\mu(x))$ defined by $(U\Psi)(x) = \Psi(x)\cdot\exp(-v_0x/2D)$ reduces our problem to the study of the differential operator $\Lambda' = -(v_0^2/4D)I + Dd_x^2$ with boundary condition $v_0\Psi(0) + 2D(d_x\Psi)(0) = 0$. The theorem follows then directly from Theorem XIII. 5. 35 of Dunford and Schwartz[7] and from Lemma 2. ■

Proof of Lemma 3. Since $S_0(t)\Phi_0 = \Phi_0$ for all $t \in \mathbb{R}^+$, it is sufficient to restrict our attention to $\mathcal{H}_0^+$. Let $\{E_\lambda | -\infty < \lambda \leq -v_0^2/4D\}$ be the spectral family associated with Λ_0^+. By definition $\Psi \in \mathcal{H}_0^+$ belongs to $\mathfrak{D}[(\Lambda_0^+)^n]$ if and only if

$$\|(\Lambda_0^+)^n\Psi\|_\mu^2 = \int_{-\infty}^{-a} \lambda^{2n} d\|E_\lambda\Psi\|_\mu^2 < \infty, \quad \text{with } a = v_0^2/4D.$$

For any $\Phi \in \mathcal{H}_0^+$ and any $t > 0$ we have

$$\|(\Lambda_0^+)^n S_0(t)\Phi\|_\mu^2 = \int_{-\infty}^{-a} \lambda^{2n}\exp(2\lambda t) d\|E_\lambda\Phi\|_\mu^2 < B_n^2\|\Phi\|_\mu^2,$$

where

$$B_n = (n/et)^n = A_n t^{-n} < \infty \quad \text{whenever } t > 0. \quad ■$$

Proof of Theorem 3: We recall that $\Phi_0(x) = (v_0/D)^{1/2}\exp(-v_0x/D)$. Clearly, $\Phi_0 \in \mathfrak{D}(X)$ and thus $\phi(t)$ is well defined. $\{S_0(t) | t \in \mathbb{R}^+\}$ is strongly continuous and uniformly bounded. By Schwartz inequality this proves (a). Since $\phi(t)$ is continuous

$$\lim_{t\to+0} \phi(t) = \phi(0) = (\Phi_0, X^2\Phi_0)_\mu$$

and a direct computation shows that $(\Phi_0, X^2\Phi_0)_\mu = 2(D/v_0)^2$. This proves the first part of (b). Let E_0 be the projector from $\mathcal{H}_0$ onto $\mathcal{H}^{(0)}{}_0$. We have then

$$|\phi(t) - (\Phi_0, XS_0(t)E_0X\Phi_0)_\mu| = |(\Phi_0, XS_0(t)(I - E_0)X\Phi_0)_\mu|$$
$$\leq \|X\Phi_0\|_\mu \cdot \|(I - E_0)X\Phi_0\|_\mu \exp(-v_0^2t/4D).$$

Hence

$$\lim_{t\to+\infty} | \phi(t) - (\Phi_0, X\Phi_0)_\mu^2 | = 0.$$

A direct computation shows that $(\Phi_0, X\Phi_0)_\mu = (D/v_0)$. This completes the proof of part (b) of the theorem. Part (c) follows directly from Lemma 3. The proof of part (d) requires some straightforward classical analysis. From the explicit form of the generalized Fourier transform used in the proof of Theorem 2, we find that

$$\phi^{(n)}(t) = (-1)^n v_0^2 (v_0^2/4D)^{n-2} \pi^{-1} \int_0^\infty dx\, x^2 (1+x^2)^{n-3} \times \exp[-v_0^2(1+x^2)t/4D].$$

In particular,

$$\lim_{t\to+0} \phi'(t) = \phi'(0) = -(4D/\pi) \int_0^\infty dx\, x^2 (1+x^2)^{-2}$$
$$= -(2D/\pi) \int_0^1 du\, u^{-1/2}(1-u)^{1/2} = -(2D/\pi) B(1/2;3/2)$$
$$= -(2D/\pi)\Gamma(1/2)\Gamma(3/2)/\Gamma(2) = -D.$$

This proves the first assertion of part (d). To prove the second, we start again (this time with $n = 2$) from the explicit form of $\phi^{(n)}(t)$ given above, and we get

$$\phi''(t) = v_0^2 (\exp - v_0^2 t/4D)\pi^{-1} \int_0^\infty dx\, x^2 (1+x^2)^{-1} \times \exp(-v_0^2 x^2 t/4D)$$
$$= v_0 D^{1/2} \exp(-v_0^2 t/4D)\pi^{-1} \times \int_0^\infty du\, u[u + (v_0^2/4D)]^{-1} \exp(-ut)$$
$$= v_0 (D/\pi)^{1/2} t^{-1/2} \exp(-v_0^2 t/4D) - v_0^2 \pi^{-1/2} \int_{a(t)}^\infty du \exp(-u^2)$$

where $a(t) = (v_0^2 t/4D)^{1/2}$.

From this follows immediately the last statement of the theorem. ∎

Theorem 5 depends on Theorem 4 only inasmuch as Theorem 4 establishes the existence of the mechanical models dealt with in Theorem 5. Since the assertions of Theorem 5 are not limited to minimal mechanical models, we prove that theorem first.

Proof of Theorem 5: To say that $(\mathfrak{A}, \{\alpha(t) \mid t \in \mathbb{R}\}, \phi)$ is a mechanical model for $(\mathcal{C}_0(\mathbb{R}^+), \{S_0(t) \mid t \in \mathbb{R}^+\}, \phi_0)$ means that for all A and B in $\mathcal{C}_0(\mathbb{R}^+)$ and all $t \in \mathbb{R}^+$ we have

$$(\Phi, AU(t)B\Phi) = (\Phi_0, AS_0(t)B\Phi_0) = (\Phi_0, B^* S^*_0(t) A^* \Phi_0)^*,$$

Since $S_0(t)$ is self-adjoint, this is equal to

$$(\Phi_0, B^* S_0(t) A^* \Phi_0)^* = (\Phi, B^* U(t) A^* \Phi)^* = (\Phi, AU(-t)B\Phi)$$

from which the first part of the theorem follows immediately. Moreover, since $\mathcal{C}_0(\mathbb{R}^+)\Phi_0$ is dense in $\mathcal{H}_0$ we obtain by continuity that for all $t \in \mathbb{R}$ and all Ψ and Ψ' in $\mathcal{H}_0$ we have $(\Psi', U(t)\Psi) = (\Psi', S_0(|t|)\Psi)$. Hence $(\Psi' U(t)\Psi)$ is an even function of t. To impose in addition that Ψ belong to $\mathcal{D}(H)$ implies that the derivative of this function exists and is continuous and thus vanishes at $t = 0$. Hence $\Psi \in \mathcal{D}(H) \cap \mathcal{H}_0$ implies that the derivative of $(\Psi', S_0(t)\Psi)$ exists and vanishes at $t = 0$, i.e., that $(\Psi', \Lambda_0 \Psi) = 0$ for all $\Psi' \in \mathcal{H}_0$, i.e., $\Lambda_0 \Psi = 0$. Since 0 is a nondegenerate eigenvalue of Λ_0, Ψ must indeed by colinear to Φ_0. This concludes the proof of Theorem 5. ∎

Proof of Theorem 4: To prove the first two parts of the theorem one could invoke the Nagy extension theorem.[16] For reasons which will become clear later on, we prefer to conduct the proof by sketching the construction of an explicit example. To this effect, consider the subspace $\mathcal{K}$ of $\mathcal{H}_0 \otimes \mathcal{L}^2(\mathbb{R}, dy)$ defined as

$$\mathcal{K} = \{\lambda \Phi_0 \otimes 1 \mid \lambda \in \mathbb{C}\} \oplus \mathcal{H}_0^+ \otimes \mathcal{L}^2(\mathbb{R}, dy),$$

where $\mathcal{H}_0 = \mathcal{L}^2(\mathbb{R}^+, \exp(v_0 x/D)dx)$, $\Phi_0(x) = (v_0/D)^{1/2} \exp(-v_0 x/D)$, $\mathcal{H}_0^+$ is the orthocomplement of Φ_0 in $\mathcal{H}_0$, and $1 \in \mathcal{L}^2(\mathbb{R}, dy)$ is the function $1(y) = \chi_{[0,\infty)}(y) \exp(-y/2)$. We identify $\mathcal{H}_0$ with the subspace $\mathcal{H}_0 \otimes 1 \subset \mathcal{K}$, and $\mathfrak{A}_0$ with $\mathcal{C}_0(\mathbb{R}^+) \otimes I$. Let $\{U(t) \mid t \in \mathbb{R}\}$ be the weakly-continuous, one-parameter group of unitary operators acting on $\mathcal{K}$, generated by the self-adjoint operator $H = \Lambda_0 \otimes (2P)$ where $P = -id_y$. We then define $\mathfrak{A}$ as the C^*-algebra generated by the elements of $\mathcal{B}(\mathcal{K})$ of the form $U(t)A_0U(-t)$ for all $t \in \mathbb{R}$ and all A_0 in $\mathfrak{A}_0$. Clearly, $U(t)AU(-t)$ generates an automorphism group $\{\alpha(t) \mid t \in \mathbb{R}\}$ of $\mathfrak{A}''$. Finally, we take for Φ the normalized vector $\Phi_0 \otimes 1$, and for ϕ the resulting state on $\mathfrak{A}$. To check that $(\mathfrak{A}, \{\alpha(t) \mid t \in \mathbb{R}\}, \phi)$ is a minimal mechanical model for $(\mathcal{C}_0(\mathbb{R}^+), \{S_0(t) \mid t \in \mathbb{R}^+\}, \phi_0)$ it is sufficient to check that the restriction $\{U^+(t) \mid t \in \mathbb{R}\}$ of $\{U(t) \mid t \in \mathbb{R}\}$ to $\mathcal{H}_0^+ \otimes \mathcal{L}^2(\mathbb{R}, dy)$ satisfies the following properties:

(i) $(\Psi_1, U^+(t)\Psi_2) = (\Psi_1, S_0^+(t)\Psi_2)$ for all $t \in \mathbb{R}^+$, and all Ψ_1, Ψ_2 in $\mathcal{H}_0^+ \otimes 1$;

(ii) $\{U^+(t)\Psi \mid t \in \mathbb{R},\ \Psi \in \mathcal{H}_0^+ \otimes 1\}$ generates a dense linear manifold in $\mathcal{H}_0^+ \otimes \mathcal{L}^2(\mathbb{R}, dy)$.

This is easily done if one first performs the unitary transformations indicated in the proof of Theorem 2 which generated an isometric isomorphism from $\mathcal{H}_0^+ \otimes \mathcal{L}^2(\mathbb{R}, dy)$ on $\mathcal{L}^2(\mathbb{R}^+ \times \mathbb{R}, dxdy)$ and transform $\{U^+(t) \mid t \in \mathbb{R}\}$ into $\{U'(t) \mid t \in \mathbb{R}\}$ with

$$(U'(t)\Psi)(x,y) = \Psi(x, y - 2\lambda(x)t) \quad \text{with } \lambda(x) = (v_0^2/4D) + Dx^2.$$

This straightforward computation concludes the proof of the first part of Theorem 4. To prove the second part of the theorem it is sufficient to notice that the pair $(\mathcal{K}, \{U(t) \mid t \in \mathbb{R}\})$ attached to a minimal mechanical model $(\mathfrak{A}, \{\alpha(t) \mid t \in \mathbb{R}\}, \phi)$ of a stationary dissipative system $(\mathfrak{A}_0, \{S_0(t) \mid t \in \mathbb{R}^+\}, \phi_0)$ is uniquely determined, up to unitary equivalence, by the pair $(\mathcal{H}_0, \{S_0(t) \mid t \in \mathbb{R}^+\})$ defined by this system. To prove the last part of the theorem, it is sufficient to prove that the spectrum of the generator H^+ of $\{U^+(t) \mid t \in \mathbb{R}\}$ is absolutely continuous with respect to Lebesgue measure, covers $\mathbb{R}$, and is countably infinitely degenerated. This is achieved by noticing that the unitary transformation

$$(V\Psi)(x,y) = (2\lambda(x))^{1/2} \Psi(x, 2\lambda(x)y)$$

from $\mathcal{L}^2(\mathbb{R}^+ \times \mathbb{R}, dxdy)$ onto itself transforms $\{U'(t) \mid t \in \mathbb{R}\}$ into $\{U''(t) \mid t \in \mathbb{R}\}$ with

$$(U''(t)\Psi)(x,y) = \psi(x, y - t) = \exp[-i(I \otimes P)t]\Psi(x,y).$$

Clearly, $I \otimes P$ has the desired spectral properties and so has thus H^{+} to which it is unitarily equivalent. ■

[1]See, for instance, G. H. Wannier, *Statistical Physics* (Wiley, New York, 1966).
[2]W. Nerst, Z. Phys. Chem. (Frankf. a. M.) **9,** 613 (1884).
[3]Five papers published between 1905 and 1908 by Einstein on diffusion and Brownian motion are collected in A. Einstein, *Investigations on the Theory of the Brownian Movement*, edited by R. Furth (A.D. Cowper, transl.) (Methuen, London, 1926).
[4]E. Nelson, *Dynamical Theories of Brownian Motion* (Princeton U. P.,Princeton, N. J., 1967).
[5]J. L. Lebowitz, in *Statistical Mechanics*, edited by S. A. Rice, K. F. Freed, and J. C. Light (University of Chicago Press, Chicago, 1972).
[6]See, for instance, Th.I.1.14 and the last lemma to Th.II.1.3 in G. G. Emch, *Algebraic Methods in Statistical Mechanics and Quantum Field Theory* (Interscience, New York, 1972). The generality with which these results can be obtained indicate how the present investigation would be extended to the discussion of quantum transport processes.
[7]N. Dunford and J. T. Schwartz, *Linear Operators* (Interscience, New York, 1963), Part II.
[8]D. W. Jepsen, J. Math. Phys. **6,** 405 (1965).
[9]O. de Pazzis, Commun. Math. Phys. **22,** 121 (1971).
[10]J. L. Lebowitz and J. Sykes, preprint (1972).
[11]V. I. Arnold and A. Avez, Ergodic Problems of Classical Mechanics (Benjamin, New York, 1968).
[12]D. A. Dubin and G. Sewell, J. Math. Phys. **11,** 2990 (1970).
[13]G. G. Emch and H. J. F. Knops, J. Math. Phys. **11,** 3008 (1970).
[14]See, for instance, L. van Hove, Physica (Utr.) **21,** 517 (1955); Physica (Utr.) **21,** 901 (1955); Physica (Utr.) **23,** 441 (1957); Physica (Utr.) **25,** 268 (1959).
[15]See for instance the version of the Ford–Kac–Mazur model presented by E. B. Davies, preprint, 1973.
[16]See B. Sz.-Nagy, Appendix in F. Riesz and B. Sz.-Nagy; *Leçons d'analyse fonctionelle* (Gauthier-Villars, Paris, 1955), 3rd ed.

Parastatistics and the quark model

A. J. Bracken* **and H. S. Green**

Department of Mathematical Physics, University of Adelaide, Adelaide, S. Australia
(Received 26 March 1973)

An extension of ordinary parastatistics is considered which makes use of all the representations of the parastatistics algebra obtained from the usual ansatz. Govorkov's demonstration that such an extension, for parastatistics of order 2, implies a $U(2)$ symmetry, is generalized for parastatistics of order p. The parastatistics algebra, restricted to N dynamical states, is characterized by the irreducible representations of $U(N)$, $SO(2N)$, and $SO(2N+1)$ which it contains. It is shown that these representations have multiplicities equal to the dimensions of associated representations of $U(p)$, $O(p)$ and $C(p)$, respectively, where $C(p)$ is a subalgebra of the enveloping algebra of $O(p)$, but is not a Lie algebra. The symmetric group $S(p)$ also appears, as a subalgebra of the enveloping algebra of $C(p)$. It is shown how a nondegenerate vacuum state may be defined for the generalized parastatistics algebra of order p, and how to construct state vectors corresponding to arbitrary numbers of quarklike particles and antiparticles. Such states belong to irreducible representations of $U(N)$, and can be obtained by the application of one kind of creation and annihilation operators to certain basic states, here called reservoir states, which correspond to the different irreducible representations of $SO(2N+1)$. The specialization to parastatistics of order 3 is discussed in detail with the application to a quark model of the hadrons in view. It is shown how to define isospin and hypercharge in a significant way in this model, which, however, differs in some respects from Gell-Mann's well-known 3-fermion model, and also from Greenberg's 3-parafermion model. Some of the physical implications are examined.

1. INTRODUCTION

The idea that all particles appearing in nature should be formed from particles of spin half is an old one, which suggested, for instance, de Broglie's theory of fusion,[1] the neutrino theory of light,[2] and Yukawa's nonlocal model for composite particles.[3] All such theories have met with grave difficulties, but a more recent manifestation of the same idea, the quark model of the hadrons,[4] has been sufficiently successful to be taken seriously. Since quarks have not been positively identified in isolation,[5] however, there is ample room for speculation concerning their nature and properties.

In the original proposal of Gell-Mann,[4] a triplet of quark fermi fields is considered, the three types of quarks and antiquarks being assigned to the triplet and antitriplet representations of $U(3)$. However, Greenberg[6] has suggested that quarks may be parafermions, satisfying parastatistics of order 3. As Fritzsch and Gell-Mann[7] have pointed out, the introduction of a triplet of parafermi fields of order 3 is in a certain sense equivalent to the introduction of nine fermi fields, together with supplementary conditions which place a restriction on the allowed states. To see this, one need only recall that a parafermi field of order 3 may be thought of as constructed from three (commuting) fermi fields via the ansatz introduced by one of us[8] in the original formulation of parafermi field quantization. The supplementary conditions then reflect the fact that the ansatz yields a reducible representation of the parafermi field algebra, from which a suitable irreducible representation is to be selected.

These observations suggest a further possibility, that only one type of parafermi field of order 3 need be introduced in order to describe all the hadrons and the associated $U(3)$ multiplet structure. In the context of the usual formulation of the quantization of such a field, this suggestion proves unrewarding, because the states available do not form complete $U(3)$ multiplets. This deficiency may be traced to the fact that one has restricted one's attention to an irreducible representation of the parafermion algebra, and, following Govorkov,[9] one is led to consider more general representations previously rejected on the grounds that the "vacuum state" appeared to be degenerate.[10] From Govorkov's point of view, this degeneracy was only a consequence of the incorrect identification of the vacuum state, and the generalized parafermi statistics had real possibilities for the definition of such physical quantities as isospin and hypercharge. He considered generalized parafermi statistics of order $p = 2$, showed that n-particle states form $U(2)$ multiplets, and gave the expressions for the $U(2)$ generators. However, he was unable to complete the corresponding task for $p = 3$, although he did indicate that one-, two-, and three-particle states form $SU(3)$ multiplets, and, despite some effort,[11] his idea has not been properly realized.

In the meantime, progress has been made in the investigation of parastatistics algebras,[12] and the authors[13] have described the structure of those representations of the parafermion algebras usually adopted, with emphasis on the representations of subalgebras isomorphic to the Lie algebras of certain unitary groups. Here we undertake a similar task for the generalized parafermion algebras, with particular reference to the case $p = 3$.

Irreducible representations of the algebra of N pairs of parafermion creation and annihilation operators are known to correspond to irreducible representations of $SO(2N+1)$. We adopt a certain reducible representation of $SO(2N+1)$, defined by the well-known ansatz,[8] and are then concerned with labeling not only the various irreducible representations of $SO(2N+1)$ which occur, but also the irreducible representations of $U(N)$ contained within each irreducible representation of $SO(2N+1)$, since these correspond to collections of states with a fixed number of particles present, and a fixed symmetry type. So we are led to consider a rather difficult state labeling problem, essentially that of *completing* the set of commuting operators provided by the invariants of the chain,

$$SO(2N+1) \supset O(2N) \supset SO(2N) \supset U(N)[\supset U(N-1) \supset \cdots \supset U(1)]. \quad (1)$$

We find that there is a related chain of algebras also represented in the space of generalized parafermi statistics of order p, viz.

$$U(p) \supset O(p) \supset SO(p) \to C(p), \tag{2}$$

where $SO(p) \to C(p)$ means that the algebra $C(p)$, which is not a Lie algebra, is a subalgebra of the enveloping algebra of $SO(p)$. Recently Drühl, Haag, and Roberts, and subsequently Ohnuki and Kamefuchi,[14] have considered the generalized parafermion algebra from a quite different point of view, and in their work, the chain $U(p) \supset O(p) \supset SO(p)$ is identified and discussed as a chain of "non-Abelian gauge groups." We find that in fact the algebras $U(p) \otimes U(N)$, $O(p) \otimes SO(2N)$, $SO(p) \otimes O(2N)$, and $C(p) \otimes SO(2N+1)$ are all represented, and the Casimir invariants of associated algebras, such as $U(p)$ and $U(N)$, are so related that the problem of completing the commuting set of operators defined by the chain (1) is precisely that of completing the set defined by the chain (2), in a certain class of representations of $U(p)$. For $p \geq 3$, this is well-known to be a very difficult problem, and we do not find an explicit solution. However, for $p = 3$ we do find an operational way of establishing a satisfactory basis in the representation space, thus implicitly defining a solution to the problem. Stated briefly, the picture which emerges in that case is as follows.

One begins with an irreducible representation of three commuting fermi fields, the ansatz components. This contains a reducible representation of the parafermion algebra of order 3, whose irreducible components may be labeled completely by the eigenvalues of two operators I and I_3, which we identify with the total isospin and third component of the isospin: Thus for each $I = 0, \frac{1}{2}, 1, \ldots$, one has $I_3 = I, I-1, \ldots, -I$. Representations labeled by different values of I are inequivalent, while those labeled by the same I and different I_3 are equivalent. The representation with $I = I_3 = 0$ is the one usually adopted for the descriptions of parafermions of order 3, and it contains the nondegenerate vacuum state. Each of the other representations contains degenerate "reservoir states", which may be thought of as containing a number $2I$ of isospin-carrying particles and antiparticles, but on which all the parafermion annihilation operators vanish. Other states within each representation are obtained by applying polynomials in the parafermion creation operators to such reservoir states. Every state in a particular basis for the Fock space of the system can be characterized by the $U(3)$ multiplet to which it belongs and its associated eigenvalues of I, I_3 and the hypercharge Y [together with the $U(N-1) \supset U(N-2) \supset \cdots \supset U(1)$ labels associated with the space–time degrees of freedom]. The $U(3)$ multiplet structure is precisely the same as that obtained in the usual model with three anticommuting fermi fields. Indeed one can assert that within the space of the generalized parafermion algebra, there acts irreducibly the algebra of three anticommuting fermi fields, in terms of which the physical $U(3)$ generators may be defined in the usual way. However, such fields have no important role in the physical interpretation of the model. For example, they carry definite quanta of Y and I_3, while the parafermion creation and annihilation operators, which are more appropriately associated with the "quarks" in this model, are isoscalars and do not carry definite quanta of hypercharge. The hypercharge and other $U(3)$ labels only become diagonal in this picture when an n-particle state is appropriately symmetrized. So we have here a model which reproduces all the multiplet structure of the usual quark model, but has a quite different interpretation at the level of the constituent subparticles.

Section 2 is concerned with analysis of the structure of the representation space for quantized parafermi statistics of order p. Emphasis is given to the representations of the algebras $SO(2N+1) \supset O(2N) \supset SO(2N) \supset U(N)$ and $U(p) \supset O(p) \supset SO(p) \to C(p)$. The labeling problem is precisely defined. In Section 3, the structure and multiplicity of "reservoir states" are examined, together with the way in which other states can be constructed from these by application of parafermion creation operators. Section 4 is devoted to a more detailed treatment of the case $p = 3$, with implicit solution for the state-labeling problem, and definition of the $U(3)$ generators. A brief description of the physical interpretation of some $U(3)$ multiplets is given.

2. GENERALIZED PARASTATISTICS

The absorption and emission operators of a kind of parafermions will be represented by a_r and $a^r = a_r^*$, and for convenience we shall suppose that the affix r takes only a finite number of values $1, 2, \ldots, N$, though in the applications N is, of course, unlimited. As in our earlier paper,[13] we define a_ρ as equal to a_r when $\rho = r$, and equal to a^r when $\rho = r + N$, so that the Greek subscript takes values from 1 to $2N$. Then, if the nonvanishing elements of $g_{\rho\sigma}$ are

$$g_{\rho\sigma} = 1, \qquad |\rho - \sigma| = N,$$

the commutation relations

$$[a_\rho, a_\sigma] = 2\alpha_{\rho\sigma},$$
$$[a_\rho, \alpha_{\sigma\tau}] = g_{\rho\sigma}a_\tau - g_{\rho\tau}a_\sigma$$

imply that the a_ρ and $\alpha_{\sigma\tau}$ may be regarded as generators of a representation of $SO(2N+1)$. We also define $g^{\rho\sigma} = g_{\rho\sigma}$, and $a^\rho = g^{\rho\sigma}a_\sigma$, etc.

If an arbitrary irreducible representation of $SO(2N+1)$ is labeled in the usual way by its highest weight $(L_1, L_2, \ldots, L_N)$, the representation $([\frac{1}{2}p]^N) = (\frac{1}{2}p, \frac{1}{2}p, \ldots, \frac{1}{2}p)$ corresponds[13] to ordinary parastatistics of order p. In the generalization which we wish to consider, the reducible representation $([\frac{1}{2}]^N) \otimes ([\frac{1}{2}]^N) \otimes \cdots \otimes ([\frac{1}{2}]^N)$, with p factors, is adopted. Then each irreducible representation with $\frac{1}{2}p \geq L_1 \geq L_2 \geq \ldots \geq L_N \geq 0$ (and $\frac{1}{2}p - L_r$ integral) occurs with a definite multiplicity in the corresponding representation space, which we shall denote by H_p. In particular, the ordinary parastatistics representation $([\frac{1}{2}p]^N)$ occurs once.[15] If $w = \frac{1}{2}p$ when p is even, but $w = \frac{1}{2}(p-1)$ when p is odd, and M_j is the number of L_r not less than j $(j = 1, 2, \ldots, w)$, the representations $(L_1, L_2, \ldots, L_N)$ may also be labeled $[M_1, M_2, \ldots, M_w]$, where $N \geq M_1 \geq M_2 \geq \cdots \geq M_w \geq 0$ (and M_j is integral).

The L_r and M_j can be regarded as operators in H_p with eigenvalues constant within any irreducible representation of $SO(2N+1)$. In this sense they are connected with the Casimir invariants of $SO(2N+1)$, constructed from the generators a_ρ and $\alpha_{\rho\sigma}$. For example, the quadratic invariant of $SO(2N+1)$ is[16]

$$\begin{aligned} C_2(2N+1) &= \alpha^\rho{}_\sigma \alpha^\sigma{}_\rho + a^\rho a_\rho \\ &= 2\sum_r L_r(L_r + 2N + 1 - 2r) \\ &= 2\Big(N(\tfrac{1}{2}p - w)(\tfrac{1}{2}p - w + N) \\ &\quad - \sum_i M_i(M_i - p + 2w + 1 - 2i - 2N)\Big). \end{aligned} \tag{3}$$

The L_r (or M_i), of course, cannot distinguish between isomorphic representations of $SO(2N+1)$ contained in H_p.

Within each representation of $SO(2N+1)$, there are generally several irreducible representations of $SO(2N)$, with generators $\alpha_{\rho\sigma}$. Any one of these is labeled by its highest weight $(\lambda_1, \lambda_2, \ldots, \lambda_N)$, which occurs once and only once within the representation $(L_1, L_2, \ldots, L_N)$ of $SO(2N+1)$, provided $L_r - \lambda_r$, $\lambda_r - L_{r+1}$ and $L_{r+1} - \lambda_{r+1}$ are nonnegative integers for all $r < N$ and $|\lambda_N| \leq L_N$. It follows that if the representation $(\lambda_1, \lambda_2, \ldots, \lambda_N)$ occurs, so does the representation $(\lambda_1, \lambda_2, \ldots, -\lambda_N)$. This pair of irreducible representations of $SO(2N)$ (or single irreducible representation in the case $\lambda_N = 0$) forms an irreducible representation of $O(2N)$, which can also be labeled by $[\mu_1, \mu_2, \ldots, \mu_w]$, where μ_j is the number of $|\lambda_r|$ not less than j $(j = 1, 2, \ldots w)$. Then $N \geq \mu_1 \geq \mu_2 \geq \cdots \geq \mu_w \geq 0$. The λ_r and μ_j are related to the Casimir invariants of $SO(2N)$ and $O(2N)$, constructed from the $\alpha_{\rho\sigma}$, with the quadratic invariant being given by[16]

$$\begin{aligned}\sigma_2(2N) &= \alpha^{\rho}{}_{\sigma}\alpha^{\sigma}{}_{\rho} \\ &= 2\sum_r \lambda_r(\lambda_r + 2N - 2r) \\ &= 2\Big(N(\tfrac{1}{2}p - w)(\tfrac{1}{2}p - w + N - 1) \\ &\quad - \sum_i \mu_i(\mu_i - p + 2w + 2 - 2i - 2N)\Big). \end{aligned} \tag{4}$$

The λ_r (or μ_j) cannot distinguish between isomorphic representations of $SO(2N)$ or $O(2N)$ contained in H_p.

Within each representation of $SO(2N)$, there are generally many irreducible representations of $U(N)$, with generators

$$a^r{}_s = \tfrac{1}{2}([a^r, a_s] + p\delta^r{}_s).$$

Any such representation is labeled by its highest weight $(l_1, l_2, \ldots, l_N)$, where [since the weights of $SO(2N+1)$ and $SO(2N)$ are sets of eigenvalues of $a^1{}_1 - \frac{1}{2}p, a^2{}_2 - \frac{1}{2}p$, etc.] $0 \leq l_{r+1} \leq l_r \leq \lambda_r + \frac{1}{2}p \leq L_r + \frac{1}{2}p \leq p$. The l_r are integers, which can be regarded as the lengths of the rows in the Young diagram associated in the usual way with the symmetry type of the corresponding tensor representation of $U(N)$. Alternatively, the representation may be labeled $[m_1, m_2, \ldots, m_p]$, where m_i is the number of l_r not less than i $(i = 1, 2, \ldots, p)$, or the length of the ith column in the associated Young diagram. Then $N \geq m_1 \geq m_2 \geq \cdots \geq m_p \geq 0$. The l_r and m_i are related to the Casimir invariants of $U(N)$, constructed from the $a^r{}_s$, with, for example,[17]

$$\begin{aligned} a^r{}_r &= \sum_r l_r = \sum_i m_i, \\ a^r{}_s a^s{}_r &= \sum_r l_r(l_r + N + 1 - 2r) \\ &= -\sum_i m_i(m_i - N + 1 - 2i). \end{aligned}$$

The l_r (or m_i) cannot distinguish between isomorphic representations of $U(N)$ contained in H_p [nor indeed between isomorphic representations of $U(N)$ contained in an irreducible representation of $SO(2N)$]. We now turn our attention to the problem of describing the multiplicity with which isomorphic representations of $U(N)$ occur in H_p.

One way of defining the generalized parastatistics representation is via the ansatz[8]

$$a_\rho = \sum_{i=1}^{p} \sigma_\rho{}^{(i)}, \tag{5}$$

in which the $\sigma_\rho{}^{(i)}$ are fermion creation and annihilation generators for a fixed value of i, but commute for different values of i. The only way we shall make use of this ansatz, however, is by the properties with which it endows the operators

$$c^{(ij)} = \sigma^{r(i)}\sigma_r{}^{(j)},$$

in particular

$$\begin{aligned} &[c^{(ij)}, a^r{}_s] = 0, \\ &\{c^{(ij)}, c^{(jk)}\} = c^{(ik)} \quad (i \neq j \neq k), \\ &[c^{(ij)}, c^{(ji)}] = c^{(ii)} - c^{(jj)} \quad (i \neq j), \\ &[c^{(ii)}, c^{(jj)}] = 0 \quad (i \neq j), \\ &[c^{(ii)}, c^{(ij)}] = c^{(ij)} \quad (i \neq j), \\ &[c^{(ij)}, c^{(kl)}] = 0 \quad (i \neq j \neq k \neq l) \end{aligned} \tag{6}$$

and

$$\begin{aligned} \sum_{i,j} c^{(ij)} &= \tfrac{1}{2}(a^\rho a_\rho - Np), \\ \sum_{i=1}^{p} c^{(ii)} &= a^r{}_r = \sum_r l_r = \sum_i m_i, \\ \sum_{i,j} c^{(ij)}c^{(ji)} &= -a^r{}_s a^s{}_r + (N+p)a^r{}_r \\ &= -\sum_r l_r(l_r - p - 2r + 1) \\ &= \sum_i m_i(m_i + p - 2i + 1). \end{aligned}$$

Now the operator $c^{(ii)}$ has integral eigenvalues, and it is evident from (6) that

$$\theta_i = \cos(\pi c^{(ii)}),$$

which has unit square, anticommutes with $c^{(ij)}$ and $c^{(ji)}$ provided $j \neq i$, and commutes with all other $c^{(kl)}$. Let us define

$$\begin{aligned} b_{jk} &= (i)^{j+k}\theta_a\theta_{a+1}\cdots\theta_b c^{(jk)}, \quad j < k, \\ b_{kj} &= -(i)^{j+k}\theta_a\theta_{a+1}\cdots\theta_b c^{(kj)}, \quad j < k, \\ b_{jj} &= c^{(jj)}, \end{aligned} \tag{7}$$

where the subscripts $a, a+1, \ldots, b$ of the θ's include all values between the odd integer a, equal to j or $j+1$, and the even integer b, equal to k or $k-1$. Thus $b_{12} = -i\theta_1\theta_2 c^{(12)}$, $b_{23} = ic^{(23)}$, and $b_{51} = \theta_1\theta_2\theta_3\theta_4 c^{(51)}$. Then it is easy to verify that

$$[b_{ij}, b_{kl}] = \delta_{jk}b_{il} - \delta_{il}b_{kj},$$

the commutation relations characteristic of $U(p)$. It follows that an irreducible representation of the b_{ij}, and of the $c^{(ij)}$ also, defines an irreducible representation of $U(p)$. Moreover, since the b_{ij} are $U(N)$ invariants, H_p carries a representation of $U(p) \otimes U(N)$, with generators $b_{ij}, a^r{}_s$. Now

$$\begin{aligned} &\sum_i b_{ii} = \sum_i c^{(ii)} = \sum_i m_i, \\ &\sum_{i,j} b_{ij}b_{ji} = \sum_{i,j} c^{(ij)}c^{(ji)} = \sum_i m_i(m_i + p - 2i + 1), \cdots \end{aligned}$$

and the expressions on the right sides of these equations are precisely those adopted by the Casimir invariants of $U(p)$, when the irreducible representations are labeled $(m_1, m_2, \ldots, m_p)$. In other words, each irreducible representation of $U(p) \otimes U(N)$ contained in H_p is labeled by the $m_1, m_2, \ldots, m_p$, which are the lengths of the rows in the Young diagram corresponding to the $U(p)$ representation $(m_1, m_2, \ldots, m_p)$, and the lengths

of the columns in the Young diagram corresponding to the associated $U(N)$ representation $[m_1, m_2, \ldots, m_p]$. Since there are no invariants of $U(N)$ which cannot be expressed in terms of the b_{ij}, no irreducible representation of $U(p) \otimes U(N)$ can occur more than once in H_p. Then it is not difficult to see that every such representation with $N \geq m_1 \geq m_2 \geq \cdots \geq m_p \geq 0$ must occur just once. We may say that the representation $[m_1, m_2, \ldots, m_p]$ of $U(N)$ occurs in H_p with a multiplicity equal to the dimension of the representation $(m_1, m_2, \ldots, m_p)$ of $U(p)$. This is one way of describing the structure of H_p.

Turning now to the question of the multiplicity with which isomorphic representations of $SO(2N)$ occur in H_p, we note that

$$\gamma^{(ij)} = c^{(ij)} + c^{(ji)} \quad (i \neq j)$$

is an $SO(2N)$ invariant and that, with one exception, all invariants of $SO(2N)$ can be constructed from the $\gamma^{(ij)}$. [The exception is the pseudoscalar $SO(2N)$ invariant associated with the sign of λ_N. More precisely, the $\gamma^{(ij)}$ are $O(2N)$ invariants, from which all invariants of $O(2N)$ can be constructed.] It follows from (6) that

$$\begin{aligned} &\{\gamma^{(ij)}, \gamma^{(jk)}\} = \gamma^{(ik)} \quad (i \neq j \neq k), \\ &[\gamma^{(ij)}, \gamma^{(kl)}] = 0 \quad (i \neq j \neq k \neq l) \end{aligned} \tag{8}$$

and, consequently,

$$(\gamma^{(ik)} \pm \gamma^{(jk)})\gamma^{(ij)} = -(\gamma^{(ij)} \pm 1)(\gamma^{(ik)} \pm \gamma^{(jk)}). \tag{9}$$

We define

$$\gamma = \sum_{j>i} \gamma^{(ij)} = \tfrac{1}{2}(a^\rho a_\rho - Np). \tag{10}$$

By subtraction of (4) from (3) we find

$$a^\rho a_\rho = 2\Big(N(\tfrac{1}{2}p - w) + \sum_i \mu_i(\mu_i - p + 2w + 2 - 2i - 2N) - \sum_i M_i(M_i - p + 2w + 1 - 2i - 2N)\Big),$$

and it follows that the eigenvalues of γ are integral. From consideration of the parastatistics algebra of order 2 contained within the entire algebra, for which γ reduces to $\gamma^{(12)}$, we infer that $\gamma^{(12)}$ has integral eigenvalues, and this conclusion is, of course, independent of the superscripts of $\gamma^{(ij)}$. Hence the commuting invariants $\gamma^{(12)}, \gamma^{(34)}, \ldots, \gamma^{(2w-1,2w)}$ of $O(2N)$ all have integral eigenvalues. If we introduce

$$\theta_{ij} = \cos(\pi\gamma^{(ij)}), \tag{11}$$

it follows further from (9) that θ_{ij} anticommutes with $\gamma^{(ik)}$ and $\gamma^{(jk)}$ when $i \neq j \neq k$. Let us define

$$\begin{aligned} \beta_{jk} &= -(i)^{j+k}\theta_{a,a+1}\cdots\theta_{b-1,b}\gamma^{(jk)}, \quad j < k, \\ \beta_{kj} &= -\beta_{jk}, \end{aligned}$$

where the subscripts $a, a+1, \ldots, b-1, b$ constitute the same sequence of integers as in (7). Then it is readily verified that

$$[\beta_{ij}, \beta_{kl}] = \delta_{ik}\beta_{jl} + \delta_{jl}\beta_{ik} - \delta_{jk}\beta_{il} - \delta_{il}\beta_{jk},$$

the commutation relations characteristics of $SO(p)$. It follows that an irreducible representation of the β_{ij}, and of the $\gamma^{(ij)}$ also, defines an irreducible representation of $SO(p)$. Since the eigenvalues of $i\beta_{jk}$ are integral, tensor but not spinor representations of $SO(p)$ occur. Indeed, a close examination reveals that $\theta_{ij} = \theta_i\theta_j$, so that

$$\beta_{ij} = b_{ji} - b_{ij},$$

and the $SO(p)$ algebra is a subalgebra of the $U(p)$ algebra already discussed. Now $U(p)$ contains $O(p)$ as well as $SO(p)$. In the present situation, $O(p)$ may be regarded as obtained from $SO(p)$ by the addition of the "reflections" θ_i. Since the β_{ij} are $O(2N)$ invariants, and the θ_i are $SO(2N)$ invariants, it follows that H_p carries representations of $SO(p) \otimes O(2N)$ and $O(p) \otimes SO(2N)$. Recalling that $U(N) \subset SO(2N) \subset O(2N)$, we may characterize the structure of the generalized parastatistics algebra by the diagram

$$\begin{array}{ccccc} O(2N) & \supset & SO(2N) & \supset & U(N) \\ & \searrow & \otimes & \swarrow & \\ U(p) & \supset & Q(p) & \supset & SO(p) \end{array}$$

If an irreducible representation of $SO(p)$ is labeled by its highest weight $(\kappa_1, \kappa_2, \ldots, \kappa_w)$, the quadratic invariant of $SO(p)$ is

$$\begin{aligned} c_2(p) &= \sum_{i,j} \beta_{ij}\beta_{ji} \\ &= 2\sum_{i>j} \gamma^{(ij)2} \\ &= 2\sum_{ij} \kappa_j(\kappa_j + p - 2j). \end{aligned} \tag{12}$$

But

$$\begin{aligned} c_2(2N) &= -2\sum_{j>i} \gamma^{(ij)2} + N(N-1)p + \tfrac{1}{2}Np^2 \\ &= 2N(\tfrac{1}{2}p - w)(\tfrac{1}{2}p - w + N - 1) \\ &\quad + 2\sum_j (N - \kappa_j)(\kappa_j + N + p - 2j), \end{aligned} \tag{13}$$

and by comparison with (4) above, we see that for p even

$$\begin{aligned} &\kappa_j = N - \mu_{w+1-j} \quad (j = 1, 2, \ldots, w-1), \\ &|\kappa_w| = N - \mu_1 \end{aligned} \tag{14}$$

and for p odd

$$\kappa_j = N - \mu_{w+1-j} \quad (j = 1, 2, \ldots, w). \tag{14'}$$

[Note that if p is even and $\kappa_w \neq 0$, then (14) implies $\mu_1 < N$, which in turn means $\lambda_N = 0$. Thus $\kappa_w\lambda_N = 0$ for p even. Moreover, (14) implies also that λ_N and κ_w cannot both be zero in that case.]

From these relationships between the invariants of $SO(p)$ and $SO(2N)$, we can deduce the $SO(p) \otimes SO(2N)$ structure of H_p. Each irreducible representation of $SO(p) \otimes SO(2N)$ in H_p is labeled $\{(\kappa_1, \kappa_2, \ldots, \kappa_w), (\lambda_1, \lambda_2, \ldots, \lambda_N)\}$, where the κ_j are integers related to the μ_j as in (14) or (14'), the μ_j being defined in terms of the $|\lambda_r|$ as before. Every such representation with

$$N \geq \kappa_1 \geq \kappa_2 \geq \cdots \geq |\kappa_w| \geq 0 \tag{15}$$

for p even, but

$$N \geq \kappa_1 \geq \kappa_2 \geq \cdots \geq \kappa_w \geq 0 \tag{15'}$$

for p odd, occurs just once in H_p. This is another way of describing the structure of that space.

A description of the $SO(p) \otimes O(2N)$ or $O(p) \otimes SO(2N)$ structure of the space is rather involved, especially for

p even. For p odd, it follows from what we have said that if the representation $\{(\kappa_1, \kappa_2, \ldots, \kappa_w), (\lambda_1, \lambda_2, \ldots, \lambda_N)\}$ occurs, so does the representation $\{(\kappa_1, \kappa_2, \ldots, \kappa_w), (\lambda_1, \lambda_2, \ldots, -\lambda_N)\}$. Each such pair forms an irreducible representation of $SO(p) \otimes O(2N)$, which we may label $\{(\kappa_1, \kappa_2, \ldots, \kappa_w), [\mu_1, \mu_2, \ldots, \mu_w]\}$. Then every such representation of $SO(p) \otimes O(2N)$ occurs just once in H_p, provided the conditions (14′) and (15′) are satisfied. On the other hand, each member of this pair forms an irreducible representation of $O(p) \otimes SO(2N)$. On the first member, the inversion operator $\zeta = \theta_1\theta_2 \cdots \theta_p$, which extends $SO(p)$ to $O(p)$, has the value

$$\zeta = \exp[i\pi(\lambda_1 + \lambda_2 + \cdots + \lambda_N + \tfrac{1}{2}Np)], \tag{16}$$

as may be seen by evaluating it on the state of highest weight with respect to $SO(2N)$. On the second member of the pair, the value of ζ is

$$\exp[i\pi(\lambda_1 + \lambda_2 + \cdots + \lambda_{N-1} - \lambda_N + \tfrac{1}{2}Np)],$$

which is opposite in sign to (16), since $2\lambda_N$ is an odd integer when p is odd. Irreducible representations of $O(p) \otimes SO(2N)$ may therefore be labeled $\{(\kappa_1, \kappa_2, \ldots, \kappa_w; \zeta), (\lambda_1, \lambda_2, \ldots, \lambda_N)\}$, and every such representation occurs just once in H_p, provided the conditions (14′), (15′), and (16) are satisfied.

Now we come to the question of the multiplicity with which isomorphic representations of $SO(2N + 1)$ occur in H_p. Let us define

$$C^{(ij)} = (\tfrac{1}{2} - \gamma^{(ij)})\theta_{ij}. \tag{17}$$

From (14) and (10) we see that

$$\sigma_2(2N + 1) = -2\sum_{j>i}(C^{(ij)2} - \tfrac{1}{4}) + Np(N + \tfrac{1}{2}p). \tag{18}$$

By considering the application of this result to parastatistics of order 2, it is evident that $C^{(12)2}$ is an $SO(2N + 1)$ invariant, and the same must be true of $C^{(ij)2}$. We can therefore resolve a_ρ into two parts:

$$a_\rho = (2\gamma^{(ij)} - 1)([\gamma^{(ij)}, a_\rho] + \{\gamma^{(ij)} - \tfrac{1}{2}, a_\rho\})$$

the first of which anticommutes, and the second of which commutes, with both $\gamma^{(ij)} - \frac{1}{2}$ and θ_{ij}. Hence $C^{(ij)}$ commutes with a_ρ, and is itself an invariant of $SO(2N + 1)$. From (8) it follows that

$$\{C^{(ij)}, C^{(jk)}\} = C^{(ik)} \quad (i \neq j \neq k),$$
$$[C^{(ij)}, C^{(kl)}] = 0 \quad (i \neq j \neq k \neq l).$$

Although these relations are the same as those satisfied by the $\gamma^{(ij)}$, the fact that the eigenvalues of the $C^{(ij)}$ are half-integral and not integral ensures very different properties. If we denote by $C(p)$ the algebra of the $C^{(ij)}$, then in view of (11) and (17), $C(p)$ is a subalgebra of the enveloping algebra of $SO(p)$, which relationship we denote by $SO(p) \to C(p)$. In general an irreducible representation of $SO(p)$ defines a reducible representation of $C(p)$. We shall show that irreducible representations of $C(p)$ in H_p may be labeled $(K_1, K_2, \ldots, K_w)$, where the K_j are related to the M_j already defined. Moreover, each irreducible representation of $C(p)$ provides a (usually reducible) representation of the symmetric group $S(p)$, i.e., $C(p) \to S(p)$. However, we have not been able, for general values of p, to set up a one-to-one correspondence between irreducible representations of $C(p)$ and irreducible representations of some Lie algebra, in the way that irreducible representations of the $c^{(ij)}$ can be associated with those of $U(p)$, and irreducible representations of the $\gamma^{(ij)}$ with those of $SO(p)$. [For $p = 3$, we shall see in Sec. 4 that it is possible to set up such a correspondence between $C(3)$ and $SU(2)$.]

Since the eigenvalues of $\gamma^{(ij)}$ are $0, \pm 1, \pm 2, \ldots$, the eigenvalues of $C^{(ij)}$ form a series $\frac{1}{2}, -\frac{3}{2}, \frac{5}{2}, -\frac{7}{2}, \ldots$, with a maximum value $(-1)^k(k + \frac{1}{2})$ in a particular irreducible representation. We shall suppose that, in an irreducible representation, the maximum eigenvalue of $|C^{(12)}|$ is $K_1 + \frac{1}{2}$ and that the maximum eigenvalue of $|C^{(2j-1\,2j)}|$ is $K_j + \frac{1}{2}$, when the $|C^{(2i-1\,2i)}|$ $(i < j)$ already have their maxima. Then the K_j $(j = 1, \ldots, w)$ may be used to label the representation. Let

$$P_{ij} = \cos[\tfrac{1}{2}\pi(C^{(ij)} - \tfrac{1}{2})].$$

Then it follows from the identities analogous to (9) satisfied by the $C^{(ij)}$ that P_{ij} commutes with $C^{(ik)} + C^{(jk)}$ but anticommutes with $C^{(ik)} - C^{(jk)}$, and hence

$$P_{ij}C^{(ik)} = C^{(jk)}P_{ij}, \quad i \neq j \neq k.$$

Since also $P_{ij}^2 = 1$, the P_{ij} provide a representation of the symmetric group $S(p)$.

To establish the relation between the K_j and the M_j, we note that the vector $|n\rangle$ corresponding to the maximum eigenvalue $(-1)^n(n + \frac{1}{2})$ of $C^{(ij)}$ satisfies

$$(C^{(ik)} - (-1)^n C^{(jk)})\ |n\rangle = 0 \quad (i \neq j \neq k),$$
$$(C^{(ik)2} + C^{(jk)2})\ |n\rangle = (n + \tfrac{1}{2})\ |n\rangle,$$

and hence compute the unique eigenvalue of the invariant

$$\sum_{j>i} C^{(ij)2} = \sum_{j=1}^{w}(K_j + \tfrac{1}{2})(K_j + p - 2j + \tfrac{1}{2})$$

within the irreducible representation considered. From this result and (18) we obtain the value

$$\sigma_2(2N + 1) = Np(N + \tfrac{1}{2}p) - 2\sum_j K_j(K_j + p - 2j + 1) \tag{19}$$

for the quadratic invariant of $SO(2N + 1)$. By comparison with (3) we see that

$$K_j = N - M_{w+1-j} \quad (j = 1, 2, \ldots, w). \tag{20}$$

Since the $C^{(ij)}$ are $SO(2N + 1)$ invariants, H_p carries a representation of $C(p) \otimes SO(2N + 1)$. The result (20) shows that *each* irreducible representation of $C(p) \otimes SO(2N + 1)$ in H_p is completely characterized by the K_j. It is easily seen that H_p contains, just once, each such representation of $C(p) \otimes SO(2N + 1)$ with

$$N \geq K_1 \geq K_2 \geq \cdots \geq K_w \geq 0.$$

We may say that the representation $[M_1, M_2, \ldots, M_w]$ of $SO(2N + 1)$ occurs in H_p with a multiplicity equal to the dimension of the representation $(K_1, K_2, \ldots, K_w)$ of $C(p)$, where $K_j = N - M_{w+1-j}$. This is a third way of describing the structure of H_p.

Since $SO(p) \to C(p) \to S(p)$ and $SO(2N + 1) \supset O(2N)$, we may enlarge the characteristic diagram to

$$\begin{array}{ccccccc} SO(2N+1) & \supset & O(2N) & \supset & SO(2N) & \supset & U(N) \\ & & & \searrow & \otimes & \swarrow & \\ & & U(p) & \supset & O(p) \supset SO(p) & \to & C(p) \to S(p) \end{array}$$

Because no representation of $U(p) \otimes U(N)$, $O(p) \otimes$

$SO(2N)$, $SO(p) \otimes O(2N)$, or $C(p) \otimes SO(2N+1)$ is contained more than once in H_p, it is possible to make certain deductions from the results described above. In particular, the representation $[m_1, m_2, \ldots, m_p]$ of $U(N)$ is contained in the representation $[M_1, M_2, \ldots, M_w]$ of $SO(2N+1)$ the same number of times as the representation $(N - M_w, N - M_{w-1}, \ldots, N - M_1)$ of $C(p)$ is contained in the representation $(m_1, m_2, \ldots, m_p)$ of $U(p)$. Similarly, if p is odd, the representation $[m_1, m_2, \ldots, m_p]$ of $U(N)$ is contained in the representation $(\lambda_1, \lambda_2, \ldots, \lambda_N)$ of $SO(2N)$, the same number of times as the representation $(\kappa_1, \kappa_2, \ldots, \kappa_w; \zeta)$ of $O(p)$ is contained in the representation $(m_1, m_2, \ldots, m_p)$ of $U(p)$, where the κ_j and ζ are related to the λ_r by (14′) and (16). But it is easily seen that $\zeta = \exp[i\pi(m_1 + m_2 + \cdots + m_p)]$ throughout the representation $(m_1, m_2, \ldots, m_p)$ of $U(p)$. Hence the representation $[m_1, m_2, \ldots, m_p]$ of $U(N)$ cannot occur in the representation $(\lambda_1, \lambda_2, \ldots, \lambda_N)$ of $SO(2N)$ unless

$$\exp[i\pi(m_1 + m_2 + \cdots + m_p)] = \exp[i\pi(\lambda_1 + \lambda_2 + \cdots + \lambda_N + \tfrac{1}{2}Np)].$$

We turn next to the problem of finding a suitable complete set of labeling operators in the space H_p. Such a set should contain the M_j (or K_j), since these characterize irreducible representations in H_p of the ordinary parafermion $[SO(2N+1)]$ algebra. It should also contain the m_i, since these label irreducible representations of $U(N)$, each of which corresponds to a collection of state vectors with a fixed member of particles present and with a definite symmetry type.[13] So it is appropriate to try and set up a $SO(2N+1) \supset O(2N) \supset SO(2N) \supset U(N)$ $[\supset U(N-1) \supset \cdots \supset U(1)]$ basis. The set of Casimir invariants of $SO(2N+1) \supset O(2N) \supset SO(2N) \supset U(N)$ is not in general a complete set of commuting $U(N)$ invariants in H_p. We have seen that these invariants are directly related to those of $U(p) \supset O(p) \supset SO(p) \rightarrow C(p)$, in associated representations. Then the problem of completing the former set by the addition of further suitable $U(N)$ invariants is precisely that of completing, in the associated representations of $U(p)$, the set of labeling operators provided by the Casimir invariants of $U(p) \supset O(p) \supset SO(p) \rightarrow C(p)$. It is well known[18] that it is extremely difficult to find an operator suitably to complete, in general representations of $U(3)$, the set of labeling operators provided by the invariants of $U(3) \supset SO(3)$, so the problems facing us here for $p \geq 3$ are formidable indeed. One complete set of commuting operators in a representation of $U(p)$ is provided by the invariants of the chain $U(p) \supset U(p-1) \supset \cdots \supset U(1)$. In the present situation, the corresponding subalgebras of the b_{ij} algebra are obtained by restricting the ranges of the subscripts i and j to the values 1 to $p-1$, then 1 to $p-2$, and so on. The corresponding orthonormal basis is quite unsuitable for our purposes, as it is a basis in which the K_j and κ_j are not diagonal. However, there must exist at least one orthonormal basis in which they *are* diagonal, and this basis must be related to the former one by a unitary transformation T. If we define $e_{ij} = T b_{ij} T^*$, then the e_{ij} will generate an equivalent set of $U(p)$ representations, and the invariants of the chain $U(p) \supset U(p-1) \supset \cdots \supset U(1)$ defined now in terms of the e_{ij} rather than the b_{ij}, will again be a complete set of commuting operators. In the corresponding orthonormal basis, the K_j and κ_j will be diagonal. If we can find the operator T, or at least the e_{ij}, and can identify the functional dependence of the K_j (in particular) on the invariants of $U(p) \supset U(p-1) \supset \cdots \supset U(1)$ (defined in terms of the e_{ij}), then we shall have a satisfactory solution to the labeling problem.

For $p = 2$, the problem may be solved more directly. We have in that case

$$b_{11} = c^{(11)}, \quad b_{12} = -i\theta_1\theta_2 c^{(12)}, \quad b_{21} = i\theta_1\theta_2 c^{(21)}, \quad b_{22} = c^{(22)}, \tag{21}$$

and while the operators $\sum_i b_{ii}$, $\sum_{i,j} b_{ij} b_{ji}$, and b_{11} do not comprise a suitable complete commuting set, the operators $\sum_i b_{ii}$, $\sum_{i,j} b_{ij} b_{ji}$, and $\rho = i(b_{12} - b_{21})$ do. For $\rho = \theta_{12}\gamma^{(12)} = \gamma \cos\pi\gamma$, whence $\gamma = \rho \cos\pi\rho$, and the $SO(2)$ and $C(2)$ labels κ_1 and K_1 are related to γ by (12), (17), (18), and (19), which yield

$$(\kappa_1)^2 = \gamma^2, \quad K_1(K_1+1) = \gamma(\gamma-1).$$

However, one can also give explicit expressions for the e_{ij} in this case, following Govorkov.[9] He found

$$\begin{aligned} e_{11} &= \tfrac{1}{2}(c^{(11)} + c^{(12)} + c^{(21)} + c^{(22)}), \\ e_{12} &= \tfrac{1}{2}(-c^{(11)} + c^{(12)} - c^{(21)} + c^{(22)}), \\ e_{21} &= \tfrac{1}{2}(-c^{(11)} - c^{(12)} + c^{(21)} + c^{(22)}), \\ e_{22} &= \tfrac{1}{2}(c^{(11)} - c^{(12)} - c^{(21)} + c^{(22)}), \end{aligned} \tag{22}$$

and the operators $\sum_i e_{ii}$, $\sum_{i,j} e_{ij} e_{ji}$, and e_{11} are a suitable complete commuting set, since $e_{11} - e_{22} = \gamma$.

In this paper, we are concerned mainly with the case $p = 3$. Rather than attempt to write down simple closed expressions for *all* the operators e_{ij} in that case, we shall define some of them explicitly, and the rest in a rather implicit, but nevertheless complete way. Essentially, our method involves the identification of all the states in a suitable $U(3) \supset U(2) \supset U(1)$ basis, which then defines a suitable set of $U(3)$ operators e_{ij}. We shall show how all states in H_p can be built up by applying creation operators to certain "vacuumlike" states, and subsequently how each state so constructed can be allotted $U(3) \supset U(2) \supset U(1)$ quantum numbers, depending on its mode of construction. So we have an "operational" definition of the required complete set of labeling operators. In the next section, we discuss the structure and multiplicity of "vacuumlike" states in H_p and the way in which other states in H_p can be constructed from them. These observations form the basis of our treatment of the case $p = 3$, given in the following section, and should be useful if a complete solution of the labeling problem for larger values of p ever becomes desirable.

3. PARTICLE AND ANTIPARTICLE STATES

We shall call an eigenvector in H_p of each of the $U(N)$ generators $a^r{}_r$ a state vector, or state, provided it belongs to an irreducible representation of $U(N)$. It will be called a basic state vector, or basic state, provided it also belongs to an irreducible representation of $SO(2N+1)$.

In a theory in which both particles and antiparticles are present, we assume that $N = 2W$ is even, and that, for $r \leq W$, a^r creates a particle, but for $r > W$, a_r creates an antiparticle. The vacuum state vector then belongs to the representation of $U(N)$ labeled (p^W, O^W) and is defined by

$$a_r|\,\rangle = 0, \quad r \leq W, \qquad a^r|\,\rangle = 0, \quad r > W, \tag{23}$$

together with the conditions

$$(c^{(ij)} - W\delta_{ij})\,|\,\rangle = 0.$$

This last condition ensures that the vacuum state is also a basic state, belonging to the representation of $SO(2N+1)$ labeled $([\frac{1}{2}p]^N)$. Other vectors are formed by applying a sequence of creation operators and invariants of $U(N)$ to the vacuum state vector; the resulting tensors, e.g., $a^q a^r c^{(12)} a_s c^{(23)} a^u |\ \rangle$ (where $s > W$), may be resolved into components belonging to irreducible representations of $U(N)$ which are, by definition, state vectors. We wish to describe in detail how this decomposition is to be effected.

The conditions (23) alone are insufficient to define the vacuum (unless $p = 1$), and there may be many basic states, which we call reservoir states, corresponding to vectors $|K'\rangle$ satisfying

$$\begin{aligned} a_r |K'\rangle &= 0, \quad r \leqslant W, \\ a^r |K'\rangle &= 0, \quad r > W. \end{aligned} \tag{24}$$

Within a given representation of $SO(2N+1)$, such states correspond to weights which are in the same equivalence class as the highest weight, i.e., their weights are obtained from the highest weight by certain permutations and changes of sign of its elements. Since all such weights are simple,[21] it follows that any representation of $U(N)$ containing a reservoir state can occur at most once within a given representation of $SO(2N+1)$. Moreover, it can be seen, again from the weights, that all such representations of $U(N)$ are contained in the same representation of $SO(2N)$, labeled $(L_1, L_2, \ldots, L_{N-1}, L_N)$ or $(L_1, L_2, \ldots, L_{N-1}, -L_N)$, according as W is even or odd. Supposing that $|K'\rangle$ belongs to an irreducible representation of $U(N)$ labeled $(l'_1, l'_2, \ldots, l'_N)$, or $[k'_1, k'_2, \ldots, k'_p]$, within the representation $(L_1, L_2, \ldots, L_N)$ of $SO(2N+1)$, we shall next determine the limitations on the values of the l'_r and k'_i.

By applying a suitable product of particle creation operators a^r to the reservoir state $|K'\rangle$, we can attain a vector of highest weight, in the same representation of $SO(2N+1)$, belonging to the representation of $U(N)$ labeled $(L_1 + \frac{1}{2}p, L_2 + \frac{1}{2}p, \ldots, L_N + \frac{1}{2}p)$. Since only W of the l'_r are changed in this process, at least W of the l'_r must have values not less than $\frac{1}{2}p$. Again, by applying a suitable product of antiparticle creation operators a_r to such a vector $|K'\rangle$ we can attain a vector of lowest weight in the same representation of $SO(2N+1)$, belonging to the representation of $U(N)$ labeled $(\frac{1}{2}p - L_N, \frac{1}{2}p - L_{N-1}, \ldots, \frac{1}{2}p - L_1)$. Since only W of the l'_r are changed in this process, at least W of the l'_r must have values not greater than $\frac{1}{2}p$. Hence, $p \geqslant l'_1 \geqslant \cdots \geqslant l'_W \geqslant \frac{1}{2}p \geqslant l'_{W+1} \geqslant \ldots \geqslant l'_N \geqslant 0$, and $N \geqslant k'_1 \geqslant \ldots \geqslant k'_w \geqslant W \geqslant k'_{p-w+1} \geqslant \ldots \geqslant k'_p \geqslant 0$, where, as before, $w = \frac{1}{2}p$ if p is even, but $w = \frac{1}{2}(p-1)$ if p is odd and in that event $k'_{w+1} = W$.

In general, the number of antiparticles in a reservoir state is different from zero. However, corresponding to any reservoir state $|K'\rangle$ there exists another reservoir state $|K\rangle$, within the same representation of $SO(2N)$, in which the antiparticles have been replaced by particles. Explicitly,

$$|K\rangle = \prod_{r'} (\alpha^{r\, r-W})\, |K'\rangle$$

where the product $\Pi_{r'}$ is over all values of r (greater than W) associated with antiparticles in $|K'\rangle$. Suppose that $|K\rangle$ belongs to the representation of $U(N)$ labeled $[W + k_1, W + k_2, \ldots, W + k_p]$, where $k_i \leqslant 0$ when $i > w$. If

$$S = \prod_{r=W+1}^{N} (a_r)^p$$

is the operator which creates antiparticles to saturate every level, then $S|K\rangle$ is a basic state, in the same representation of $SO(2N+1)$ as the reservoir states $|K\rangle$ and $|K'\rangle$, and is labeled $[k_1, k_2, \ldots, k_p]$. Since k_i must here be nonnegative,

$$k_i = 0, \quad i = w+1, \ldots, p.$$

By multiplying $S|K\rangle$ with a contravariant tensor operator formed entirely from the a^r (i.e., without factors a_r or $c^{(ij)}$) with a symmetry corresponding to the irreducible representation $[j_1, j_2, \ldots, j_p]$ of $U(N)$, we obtain a vector $|J, K\rangle$ of the direct product of the irreducible representations $[j_1, j_2, \ldots, j_p]$ and $[k_1, k_2, \ldots, k_p]$. This vector $|J, K\rangle$ can be resolved into basic vectors $|M(J, K)\rangle$ belonging to irreducible representations $[m_1, m_2, \ldots, m_p]$ of $U(N)$, where[17]

$$\begin{aligned} &\max(j_1 + k_p, k_1 + j_p) \leqslant m_1 \leqslant j_1 + k_1, \\ &\max(j_1 + j_2 + k_p + k_{p-1}, k_1 + k_2 + j_p + j_{p-1}) \\ &\qquad \leqslant m_1 + m_2 \leqslant j_1 + j_2 + k_1 + k_2, \\ &\qquad \cdots \\ &m_1 + m_2 + \cdots + m_p = j_1 + k_1 + j_2 + k_2 + \cdots + j_p + k_p . \end{aligned} \tag{25}$$

Each of the vectors $|M(J, K)\rangle$ belongs to the same irreducible representation of $SO(2N+1)$ as $|K\rangle$.

The above discussion suggests that isomorphic representations of $U(N)$ in H_p may be distinguished by the associated sets of values of the j_i and the k_i, and the eigenvalues of the commuting invariants of $SO(2N+1)$ (constructed from the $C^{(ij)}$) which resolve the multiplicity of representations with the same k_i, among the reservoir states. In the next section, we shall confirm that this is so for $p = 3$. However, it should be pointed out that we have no guarantee that every set of values of the j_i and k_i consistent with a given set of values of the m_i will correspond to a different isomorphic representation of $U(N)$ in general. In fact, it is easy to confirm that the only completely symmetric tensor involving p operators $a^r, a^s, \ldots, a^x$ is $\{a^r, \{a^s, \{\ldots, a^x\}\ldots\}\}$, and even in expressions where other creation and annihilation operators are present, such a symmetrized product can be separated in this form. *Hence it may be assumed that* $j_p = m_p$.

4. PARASTATISTICS OF ORDER 3

According to what has been said in Sec. 2, the representation $[m_1, m_2, m_3]$ of $U(N)$ occurs in H_3 a number of times equal to the dimension of the representation (m_1, m_2, m_3) of $U(3)$, and there act in H_3 a corresponding set of $U(3)$ generators e_{ij} $(i, j = 1, 2, 3)$ such that the Casimir invariants of the chain $U(3) \supset U(2) \supset U(1)$ form a complete set of commuting invariants of $U(N)$. Moreover, the elements of this set may be assumed to commute also with the operators K_1 (= K below) and κ_1 (= κ below), which label representations of $C(3) \otimes SO(2N+1)$ and $SO(3) \otimes O(2N)$ in H_3, so that K and κ are functions of the elements of that set.

Rather than the notation e_{ij} for all the $U(3)$ generators, we shall use the more familiar $I_3 = \frac{1}{2}(e_{11} - e_{22})$, $I_+ = I_1 + iI_2 = e_{12}$, $I_- = I_1 - iI_2 = e_{21}$, and $Y = (e_{11} + e_{22} - 2e_{33})/3$, together with e_{13}, e_{31}, e_{23}, and e_{32}. We seek to identify or characterize these operators, and to express K (in particular) as a function of the Casimir invariants of the chain $U(3) \supset U(2) \supset U(1)$, i.e., of m_1, m_2, m_3, Y, I, and I_3 (where $I^2 = I(I+1)$, $I \geqslant 0$).

By considering the reduction of the representation $([\frac{1}{2}]^N) \otimes ([\frac{1}{2}]^N) \otimes ([\frac{1}{2}]^N)$, we find that the irreducible representations of $SO(2N+1)$ contained in H_3 are labeled $([\frac{3}{2}]^{N-K}, [\frac{1}{2}]^K)$, or alternatively $[N-K]$, where $K = 0, 1, \ldots, N$, and there are $(K+1)$ isomorphs corresponding to a particular value of K. These may be distinguished by the eigenvalues $\frac{1}{2}, -\frac{3}{2}, \ldots (-1)^K(K+\frac{1}{2})$ of $C^{(12)}$. The corresponding representation of $C(3)$ is thus $(K+1)$-dimensional, and is labeled (K). Each such irreducible representation of $SO(2N+1)$ decomposes with respect to $SO(2N)$ as $([\frac{3}{2}]^{N-K-1}, [\frac{1}{2}]^K, \pm\frac{1}{2}) \oplus ([\frac{3}{2}]^{N-K}, [\frac{1}{2}]^{K-1}, \pm\frac{1}{2})$, with the exception of $([\frac{3}{2}]^N)$, which gives $([\frac{3}{2}]^{N-1}, \pm\frac{1}{2}) \oplus ([\frac{3}{2}]^{N-1}, \pm\frac{3}{2})$, and $([\frac{1}{2}]^N)$, which gives $([\frac{1}{2}]^{N-1}, \pm\frac{1}{2})$ only. Thus in H_3 there are $(2\kappa+1)$ irreducible representations of $SO(2N)$ labeled $([\frac{3}{2}]^{N-\kappa}, [\frac{1}{2}]^\kappa)$, where $\kappa = 0, 1, \ldots, N$ and a similar number with the sign of the last weight reversed; alternatively there are $(2\kappa+1)$ irreducible representations of $O(2N)$ labeled $[N-\kappa]$. The isomorphs may be distinguished by the eigenvalues $0, \pm 1, \ldots, \pm\kappa$ of $\gamma^{(12)}$, and the corresponding representation of $SO(3)$ is labeled (κ), and is $(2\kappa+1)$-dimensional. Clearly, when κ has a value κ_0, then $K = \kappa_0$ or $\kappa_0 - 1$, except that $K = 0$ if $\kappa = 0$. In other words, the $(2\kappa_0+1)$-dimensional representation $(\kappa = K_0)$ of $SO(3)$ reduces into irreducible representations of $C(3)$ of dimension κ_0, labeled $(K = \kappa_0 - 1)$, and $\kappa_0 + 1$, labeled $(K = \kappa_0)$; except that the representation (0) of $SO(3)$ yields only the representation (0) of $C(3)$. Now it is easily seen that if a representation of $U(3)$ contains representations of $SO(3)$ labeled $(\kappa_1), (\kappa_2), \ldots$, of dimension $(2\kappa_1+1), (2\kappa_2+1), \ldots$, then it contains representations of $SU(2)$ corresponding to $I = \frac{1}{2}\kappa_1, \frac{1}{2}(\kappa_1 - 1); \frac{1}{2}\kappa_2, \frac{1}{2}(\kappa_2 - 1); \cdots$, of dimension $\kappa_1 + 1, \kappa_1; \kappa_2 + 1, \kappa_2; \cdots$. It follows that the irreducible representation (K) of $C(3)$ appears in any given representation of $U(3)$ the same number of times as the irreducible representation of $SU(2)$ with $I = \frac{1}{2}K$, which has the same dimension $(K+1)$. We may therefore choose the e_{ij} in such a way that I is *equal* to $\frac{1}{2}K$, and the enveloping algebras of $C(3)$ and $SU(2)$ (with generators I_1, I_2, and I_3) are the same. Then we see from (18) and (19) that

$$4I^2 = K(K+2) = C^{(12)2} + C^{(23)2} + C^{(31)2} - \tfrac{3}{4}. \quad (26)$$

It also follows from this identification of $C(3)$ and $SU(2)$ representations that the isospin generators I_1, I_2, and I_3 are $SO(2N+1)$ invariants, commuting with all a^r and a_r. Thus these parafermion operators are associated with isoscalar particles. Note that the functional dependence of K on m_1, m_2, m_3, Y, I, and I_3 has now been fixed as simply $K = 2I$. The most natural characterization of the isospin in this formalism is by the $C(3)$ operators $C_1 = C^{(23)}$, $C_2 = C^{(31)}$, and $C_3 = C^{(12)}$, which satisfy

$$\{C_1, C_2\} = C_3 \quad (27)$$

etc., rather than by the components of **I**, which are generators of $SU(2)$. However, we shall show later in this section how the two sets of operators are related.

Next we come to the identification of the hypercharge operator Y. It is convenient to introduce the $O(2N)$ invariant γ defined in (10), which reduces here to

$$\gamma = \tfrac{1}{2}a^\rho a_\rho - 3N/2$$
$$= \gamma^{(12)} + \gamma^{(13)} + \gamma^{(23)} \quad (28)$$

and which may be seen from (13), (18), and (19) to have the eigenvalues $(K+2)$ or $-K$ in the representation $[N-K]$ of $SO(2N+1)$, and $(\kappa+1)$ or $-\kappa$ in the representation $[N-\kappa]$ of $O(2N)$. Hence $K = \kappa - 1$ if the eigenvalue of γ is positive and otherwise $K = \kappa$. Consider now the reservoir states, which, according to the analysis of the preceding section, belong to $U(N)$ representations of the form $[W+k_1, W, W+k_3]$. Here k_1 is the number of particles, and $-k_3$ the number of antiparticles in the reservoir. From (24) and (28) we find that the corresponding eigenvalues of γ are $-(k_1 - k_3)$. Thus $I = \frac{1}{2}(k_1 - k_3)$ on such reservoir states. A general M representation of $U(N)$, labeled $[m_1, m_2, m_3]$, may be regarded as belonging to the decomposition of the direct product of a J representation $[j_1, j_2, j_3]$, associated with an appropriately symmetrized product of operators a^r, and a K representation $[k, 0, 0]$, to a state of which that product is applied. Such basic states in $[k, 0, 0]$ are obtained by adding all possible antiparticles to reservoir states $|K\rangle$, which contain k particles only, and are therefore labeled $[W+k, W, W]$. (The "preceding" vectors of Govorkov[9] belong to representations labeled $[k, 0, 0]$.)

Bearing in mind that I commutes with a_ρ, one sees that if a particular M representation is associated with a certain value of I, then the corresponding K representation has $k = 2I$. Moreover, in view of the inequalities (25), the maximum isospin in a set of isomorphic M representations will be $\frac{1}{2}(m_1 - m_3)$. This can be resolved into two Casimir operators $\frac{1}{2}(m_1 - m_2)$ and $\frac{1}{2}(m_2 - m_3)$, corresponding to the isospins in the submultiplets of greatest and least hypercharge, respectively. The hypercharge Y itself should vary between a minimum value $(m_2 + m_3 - 2m_1)/3$ attained when $k = m_2 - m_3$, $j_2 = j_3$ and $j_1 = m_1$, and a maximum value $(m_1 + m_2 - 2m_3)/3$ attained when $k = m_1 - m_2, j_1 = j_2$, and $j_3 = m_3$. Thus

$$Y = \tfrac{1}{3}(m_1 + m_2 + m_3) + j_2 - j_1 - j_3.$$

As we pointed out at the end of the last section, multiplets corresponding to different values of the j_i are not always independent, and for $p = 3$ the ambiguity is most simply removed by imposing the condition

$$j_3 = m_3 \quad (29)$$

so that the formula for the hypercharge and isospin may be written

$$Y = \tfrac{1}{3}(m_1 + m_2 - 2m_3) + j_2 - j_1,$$
$$I = \tfrac{1}{2}k = \tfrac{1}{2}(m_1 + m_2) - \tfrac{1}{2}(j_1 + j_2). \quad (30)$$

Within a given representation $[m_1, m_2, m_3]$ of $U(N)$, it is evident that the values of the j_i are completely determined by Y and I, when the condition (29) is adopted; moreover, the values of Y and I allowed are just those which occur in the $U(3)$ multiplet (m_1, m_2, m_3). Thus, we have verified that the condition (29) does not exclude any states contained within the representations of the generalized parastatistics algebras.

The operators m_1, m_2, m_3, Y, I, and I_3 (which we have not yet defined) form a complete set of commuting $U(N)$ invariants in H_3, and by fixing their eigenvalues on a set of basic states we implicitly define the e_{ij} completely, since all their matrix elements are then determined. Although these $U(3)$ generators do not provide the simplest characterization of the algebraic structure—we have already seen that $C(3)$ arises more naturally than $SU(2)$—we wish to show how they can be constructed if required.

The choice of the isospin $SU(2)$ generators $I_\pm$ and I_3 is not, of course, unique, but can be made so by requiring that

$$UI_+U^* = \omega I_+, \quad UI_-U^* = \omega^2 I_-, \quad UI_3U^* = I_3, \tag{31}$$

where ω is a complex cube root of 1, and U is the operator inducing a cyclic transformation of the ansatz components. Thus

$$Ua_\rho' U^* = \omega a_\rho', \quad Ua_\rho'' U^* = \omega^2 a_\rho'', \quad Ua_\rho U^* = a_\rho, \tag{32}$$

where we have defined

$$a_\rho' = \sigma_\rho^{(1)} + \omega\sigma_\rho^{(2)} + \omega^2\sigma_\rho^{(3)},$$
$$a_\rho'' = \sigma_\rho^{(1)} + \omega^2\sigma_\rho^{(2)} + \omega\sigma_\rho^{(3)}. \tag{33}$$

in terms of the psuedofermion operators appearing in the ansatz (5). This U is in the symmetric group $S(3)$ discussed in Sec. 2. Indeed, if

$$U_j = \cos[\tfrac{1}{2}\pi(C_j - \tfrac{1}{2})],$$

it follows from (27) that, e.g., $U_3C_1 = C_2U_3$ and $U_3C_2 = C_1U_3$, and since U must be an $SO(2N+1)$ invariant, we have

$$U = U_1U_2 = U_2U_3 = U_3U_1.$$

We can easily construct one set of $SU(2)$ generators $H_\pm$ and H_3 by writing

$$H_3 = \tfrac{1}{2}[C_3 - \tfrac{1}{2}(-1)^K],$$
$$H_+ = f(C_3)(C_1 + C_2)(C_2 - C_1),$$
$$H_- = (C_2 - C_1)(C_1 + C_2)f(C_3).$$

Since

$$[H_3, H_\pm] = \pm H_\pm,$$

$$H_+H_- = [f(C_3)]^2[(K+1)^2 - (C_3 + \tfrac{1}{2})^2][(K+1)^2 - (C_3 - \tfrac{3}{2})^2],$$

$$H_-H_+ = [f(C_3 + 1)]^2[(K+1)^2 - (C_3 + \tfrac{3}{2})^2][(K+1)^2 - (C_3 - \tfrac{1}{2})^2],$$

the required commutation relations will be satisfied, provided $f(C_3)$ is defined by

$$4[f(C_3)]^2[(K+1)^2 - (C_3 + \tfrac{1}{2})^2][(K+1)^2 - (C_3 - \tfrac{3}{2})^2] = (K+1)^2 - [C_3 - \tfrac{1}{2}(-1)^K - \tfrac{1}{2}]^2.$$

It is important to note that any $SO(2N)$ vector, i.e., a linear combination of a_ρ, a_ρ', and a_ρ'' with coefficients which may be $SO(2N)$ invariants, can change the eigenvalue of H_3 by at most ± 1; the same will apply to I_3, as defined below.

The H_j and I_j are evidently connected by a unitary transformation, which we next determine. If

$$V_3 = 1 + Uu_3^* + U^*u_3,$$
$$u_3 = \exp(4\pi i H_3/3),$$

it is easy to verify that the relations (31) are satisfied, provided

$$V_3H_j = I_jV_3$$

and hence that

$$I_j = VH_jV^*,$$
$$V = \lim_{\epsilon\to 0}(V_3 + \epsilon)[(V_3^* + \epsilon)(V_3 + \epsilon)]^{-1/2}.$$

The need to take a limit here arises from the fact that V_3 has one accidentally vanishing eigenvalue, in the representation of $C(3)$ corresponding to $K = 2$. In any irreducible representation of $SO(2N+1)$, the matrix elements of U can readily constructed from the normalized eigenvectors of H_1 in a representation in which H_3 is diagonal, and those of V can be derived therefrom.

The hypercharge changing operators e_{23}, e_{32}, e_{13}, and e_{31} can be split into two parts, one of which increases the isospin I by one-half unit, and the other decreases it by one-half unit; thus

$$e_{ij} = (e_{ij})^+ + (e_{ij})^-,$$
$$(e_{ij})^+ I = (I - \tfrac{1}{2})(e_{ij})^+ \quad (i < j = 3 \text{ or } j < i = 3)$$
$$(e_{ij})^- I = (I + \tfrac{1}{2})(e_{ij})^-.$$

Instead of $(e_{ij})^+$ and $(e_{ij})^-$, we shall first construct operators $(D_{ij})^+$ and $(D_{ij})^-$ which differ from them only in normalization. We shall need to make use of the $U(N)$ invariants m_1, m_2, and m_3, and since our object is to construct all the $SU(3)$ generators at least implicitly from the creation and annihilation operators, we note that the m_j are determined by $m_1 + m_2 + m_3 = a^r{}_r$,

$$2(m_1 + 2m_2 + 3m_3) - (m_1^2 + m_2^2 + m_3^2) = a^r{}_s a^s{}_r - (N-1)a^r{}_r,$$

$$3(m_1 + 4m_2 + 9m_3) - 3(m_1^2 + 2m_2^2 + 3m_3^2) + (m_1^3 + m_2^3 + m_3^3) = a^r{}_s a^s{}_t a^t{}_r - (2N - \tfrac{3}{2})a^r{}_s a^s{}_r + \tfrac{1}{2}a^r{}_r[a^s{}_s + (N-1)(2N-1)]$$

(summation over repeated affixes implied). The operators $a^r{}_{,i} = [m_i, a^r]$ and $a_{r,i} = [a_r, m_i]$ change the eigenvalues of m_i by $+1$ and -1, respectively, leaving the other m_j unchanged. Moreover, they have no effect on I or I_3, since they commute with the $SO(2N+1)$ invariants.

The hypercharge changing operators are $U(N)$ invariants which change the value of both I_3 and I by one-half unit, and must therefore involve a_ρ' and a_ρ''. It follows from (31) and (32) that a_ρ' has components which change I_3 by $+\frac{1}{2}$ and -1, while a_ρ'' has components which change I_3 by $-\frac{1}{2}$ and $+1$. To separate the components which change I_3 by $\pm\frac{1}{2}$, we make use of the identity

$$[A^2, [A^2, [A^2, a_\rho']]] = [2A^4 - A^2, a_\rho'],$$

where we have set $A = \gamma - \frac{1}{2}$. This identity can be verified directly, or deduced from the fact that a_ρ' can have only components which change the eigenvalue A' of A to $\pm A', \pm(A'+1)$ or $\pm(A'-1)$. It is evident from (19) and (28) that the components which change I by $\pm\frac{1}{2}$ are those which change A' to $-A'$ or $A' \pm 1$, and are therefore contained in the vector

$$\alpha_\rho' = \{A, [A^2, a_\rho']\} - [A, a_\rho'].$$

Of course α_ρ'' is defined similarly in terms of a_ρ''. The vectors $\alpha'^r, \alpha''^r (r \le W)$ and $\alpha_r', \alpha_r'' (r > W)$ can be used to create particles and antiparticles respectively in the reservoir.

We may now assert that the invariants which increase the hypercharge and change the isospin (I_3, I) by fixed amounts are

$$(D_{13})^+ = [\alpha'^r, a_{r,1}], \quad (D_{23})^+ = [\alpha''^r, a_{r,1}],$$
$$(D_{13})^- = [a^r{}_{,2}, \alpha'_r], \quad (D_{23})^- = [a^r{}_{,2}, \alpha''_r].$$

For these invariants do not alter the values of m_1, m_2, and m_3, and whereas the factors $a_{\rho,i}$ change the values of j_1 and j_2, the factors α'_ρ and α''_ρ cannot. By inspection of the expression (30) for Y, it is clear that the above operators will all increase Y by one unit; their Hermitian conjugates $(D_{31})^-$, $(D_{32})^-$, $(D_{31})^+$, and $(D_{32})^+$ will similarly decrease Y by one unit. Hence we may write

$$(e_{ij})^+ = F_{ij}[(D_{ij})^+(D_{ji})^-]^{-1/2}(D_{ij})^+$$

for $i < j = 3$ and $j < i = 3$, where the F_{ij} are normalization factors known from the work of Baird and Biedenharn[19]:

$$F_{13} = F(Y, -I-1, -I_3),$$
$$F_{23} = F(Y, -I-1, I_3),$$
$$F_{31} = F(Y+1, I-\tfrac{1}{2}, -I_3+\tfrac{1}{2}),$$
$$F_{32} = F(Y+1, I-\tfrac{1}{2}, I_3-\tfrac{1}{2}),$$
$$2(I+1)(2I+1)[F(Y,I,I_3)]^2 = -(I+I_3+1)$$
$$\times (m_1 - \mu + I - \tfrac{1}{2}Y + 2)(m_2 - \mu + I - \tfrac{1}{2}Y + 1)$$
$$\times (m_3 - \mu + I - \tfrac{1}{2}Y),$$
$$\mu = \tfrac{1}{3}(m_1 + m_2 + m_3).$$

This completes the determination of the $U(3)$ generators.

Turning now to the physical interpretation implied by the above identifications, we note that there are similarities to Gell-Mann's well-known theory,[4] but also important differences. It is a requirement of Gell-Mann's theory that the fundamental particles are quarks, each of which has a definite isospin and hypercharge and, on account of its fractional charge, cannot be positively identified with any particle so far observed in nature. The generalized parastatistics also requires the hadrons to be composite particles, but the fundamental particles do not carry a definite isospin and hypercharge. The reservoir particles carry the isospin, but are of two kinds, one of which has $I_3 = -1$ or $+\frac{1}{2}$, and the other has $I_3 = -\frac{1}{2}$ or $+1$. The external particles have zero isospin, but, as can be seen from (30), have $Y = -\frac{2}{3}$ or $+\frac{4}{3}$. The indeterminacy of these quantities, and the charge $I_3 + \frac{1}{2}Y$, is resolved by the symmetry type of the state in which the particles appear. There are quarkish states with fractional hypercharge, as in Gell-Mann's theory, which can, however, be excluded by requiring that the particle number should be a multiple of 3. There is also a requirement that strong interactions should involve only the $U(N)$ generators $a^r{}_s$, which conserve isospin and hypercharge and, as Gray[20] has shown, are consistent with the cluster property which is indispensable in a theory of composite particles.

In Table I, we list the well-known hadrons and the corresponding quantum numbers suggested by the present interpretation.

The χ is of course formed by creating an antiparticle and filling the "hole" with a particle. Even excluding the quarkish states, there are obviously some simple assignments of quantum numbers to which no known stable particle can be found to correspond, notably the fermion singlet $m_1 = m_2 = m_3 = W + 1$, which can, however, be identified as a combination of a baryon and meson. If we denote the number of objects of this kind

TABLE I. Table of hadrons and assignments of quantum numbers.

Particle(s)	$m_1 - W$	$m_1 - m_2$	$m_2 - m_3$	$m_1 - j_1$	$m_2 - j_2$	I	Y
χ	0	0	0	0	0	0	0
$(\Lambda, \eta, \bar{\Lambda})$	(2, 1, 0)	1	1	0	0	0	0
$(N, K, \bar{\Xi})$	"	"	"	1	0	$\frac{1}{2}$	1
$(\Xi, \bar{K}, \bar{N})$	"	"	"	0	1	$\frac{1}{2}$	-1
$(\Sigma, \pi, \bar{\Sigma})$	"	"	"	1	1	1	0
Ω	3	3	0	0	0	0	-2
Ξ^*	"	"	"	1	0	$\frac{1}{2}$	-1
Σ^*	"	"	"	2	0	1	0
N^*	"	"	"	3	0	$\frac{3}{2}$	1

by N_1, the numbers of baryon, meson, and antibaryon octet states by N_8, N'_8, and $\bar{N}_8$, we have in general

$$m_1 - W = N_1 - \bar{N}_1 + N'_8 + 2N_8,$$
$$m_2 - W = N_1 - \bar{N}_1 + N_8 - \bar{N}_8,$$
$$m_3 - W = N_1 - \bar{N}_1 - N'_8 - 2\bar{N}_8.$$

All possible values of the m_i for which $m_1 + m_2 + m_3$ is a multiple of 3 can be obtained by suitable substitutions in this formula. It may be noticed that the higher admissible $SU(3)$ multiplets (e.g., the 27-et) can be constructed from octets, and even the decuplet can be constructed in this way.

ACKNOWLEDGMENTS

The authors are indebted to D. A. Gray for several discussions on the subject of this paper. One of us (A. J. B.) gratefully acknowledges the support provided by a Rothmans Junior Fellowship.

* *Present address:* Mathematics Department, University of Queensland, St. Lucia, Qld., Australia.

[1] L. de Broglie, *Théorie générale des particules à spin: Méthode de fusion* (Gauthier-Villars, Paris, 1943).
[2] P. Jordan, Z. Phys. **93**, 464 (1935); Z. Phys. **105**, 114 (1937); Z. Phys. **105**, 229 (1937); M. H. L. Pryce, Proc. R. Soc. A **165**, 247 (1938).
[3] H. Yukawa, Phys. Rev. **77**, 219 (1950); Phys. Rev. **80**, 1047 (1950).
[4] M. Gell-Mann, Phys. Lett. **8**, 214 (1964); J. J. J. Kokkedee (Ed.), *The Quark Model* (Benjamin, New York, 1969).
[5] See, for example, M. F. Crouch, K. Mori, and G. R. Smith, Phys. Rev. D **5**, 2667 (1972).
[6] O. W. Greenberg, Phys. Rev. Lett. **13**, 598 (1964).
[7] H. Fritzsch and M. Gell-Mann, "Light Cone Current Algebra," Tel-Aviv International Conference on Duality and Symmetry, 1971.
[8] H. S. Green, Phys. Rev. **90**, 270 (1953).
[9] A. B. Govorkov, Zh. Eksp. Teor. Fiz. **54**, 1785 (1968) [Sov. Phys.-JETP **27**, 960 (1968)]; Ann. Phys. (N.Y.) **53**, 349 (1969).
[10] O. W. Greenberg and A. M. L. Messiah, Phys. Rev. **138**, B1155 (1965).
[11] A. Ramakrishnan, R. Vasuderan, and P. S. Chandrasekaran, J. Math. Anal. Appl. 35, 249 (1971). *Added in proof:* The reader's attention is drawn to a more recent paper by A. B. Govorkov, Int. J. Theor. Phys. 7, 49 (1973).
[12] Y. Ohnuki and S. Kamefuchi, Phys. Rev. **170**, 1279 (1968); Ann. Phys. (N.Y.) **51**, 337 (1969); Ann. Phys. (N.Y.) **57**, 543 (1970); Ann. Phys. (N.Y.) **65**, 19 (1971).
[13] A. J. Bracken and H. S. Green, Nuovo Cimento A **9**, 349 (1972).
[14] K. Drühl, R. Haag and J. E. Roberts, Commun. Math. Phys. **18**, 204 (1970); Y. Ohnuki and S. Kamefuchi, "The Statistical Quantum Number and Gauge Groups in Parafermi Field Theory," Tokyo University Preprint TUEP-72-8 (1972).
[15] C. Ryan and E. C. G. Sudarshan, Nucl. Phys. **47**, 207 (1963).
[16] A. J. Bracken and H. S. Green, J. Math. Phys. **12**, 2099 (1971).
[17] H. S. Green, J. Math. Phys. **12**, 2106 (1971).
[18] G. Racah, in *Group Theoretical Concepts and Methods in Elementary Particle Physics, Lectures*, edited by F. Gürsey (Gordon&Breach, New York, 1964).
[19] G. E. Baird and L. C. Biedenharn, J. Math. Phys. **4**, 1449 (1963).
[20] D. A. Gray, "Cluster Restrictions on Parafermi Operators," Adelaide University preprint (1972).
[21] *Added in proof:* This statement is incorrect. However, the ensuing arguments are unaffected. See T. Palev, "Vacuum-Like State Analysis of the Representations of the Para-Fermi Operators', CERN Preprint TH-1653 (1973).

Generalizing Riemannian geometry

Richard H. Hudgin

Department of Physics, State University of Buffalo, Buffalo, New York
(Received 24 November 1971; final revised manuscript received 1 November 1972)

The properties of Riemannian geometry necessary to relativity have been used as a basis to derive a more general geometry. Emphasis is placed on a natural development with the result of considerable generalization. Several examples are discussed including the Brans–Dicke field equation which are but one special case of the new manifolds. The scalar field is not introduced ad hoc but is a natural geometrical quantity.

I. INTRODUCTION

Riemannian geometry has become a tool of tremendous value to the physicist and finds application in fields as different as relativity and elasticity. It is particularly the relativist, however, who has most benefitted, and it is he who has oftimes tried to generalize Riemannian geometry to gain even more insight.

This paper approaches the question of generalizing the geometry somewhat differently from past efforts.[1] Those properties which seem physically desirable are assumed, and the resulting structure is then deduced. The geometries which emerge are considerably more general than Riemannian geometry, and, in fact, even the coordinate system has a rather unusual secondary role. Several concrete examples are discussed including a derivation of the Brans–Dicke field equations as a special case.[2]

II. PHYSICAL ASSUMPTIONS

It is always a matter of judgment to extract those parts of a theory which are physically important. The following properties of Riemannian manifolds, however, have been of great use and will be constructed into the geometry.

A1. Paths
A2. Tangent vectors
A3. Metric
A4. Covariant differentiation
A5. Contraction of indicies
A6. Length preserving parallel propogation
A7. Vanishing torsion
A8. Curvature tensor
A9. Bianchi identities

What has *not* been deemed fundamental is the standard coordinate patch presentation.

III. SPACE SET

The axiomatization is begun very differently from that of a Riemannian manifold. Instead of beginning with a Hausdorffspace and coordinate patches, a net of parametrized paths is used. These paths fill the space and might be expected to be a sufficient substitute for the coordinate system.

Axiom 1: There exists an index set I called the path index set. $G = I \otimes R$ (R = set of real numbers) is called the path set. There is an assumed equivalence relation for G called a coincidence equivalence. The quotient space S is called the space set.

I is a set of labels for the paths, and G is the set of paths. Intuitively each path has a parameter t, and (i, t) is meant to represent the position (as a function of t) of the path labeled i. Since paths may intersect, each point in space may have many labels, i.e., (i, t_0) and (j, u_0) could both be the same point. Thus an equivalence relation (that of coincident points) is assumed. The resulting quotient space S is meant to be a generalization of coordinate space. Throughout this paper, whenever the meaning is clear from context, we will follow the not unusual convention where the same symbol is used for an element in G and its equivalence class in S.

IV. TANGENT VECTORS

The next step is to introduce tangent vectors. Often a tangent vector has been identified as an equivalence class of paths passing through some fixed point. In this case a function acts on a tangent vector to produce the directional derivative—which is just the derivative of the function along any of the paths in the tangent vector's equivalence class.

Definition 1: Given $p \in S$ and its equivalence class $\{p\} \subseteq G$, $T_p = R \otimes \{p\}$ is called the basic tangent space at p. An element of T_p is called a basic tangent vector with the real number its length.

Here a basic tangent vector in T_p is identified with a path i passing through p and a length α. It is thus an ordered triplet $\xi = (\alpha, i, t_0)$ where $p = (i, t_0)$.

It should be pointed out that T_p need not be all the usual tangent space. T_p might include only tangents in the open light cone or even just a finite number of tangent directions. Examples of this type will be considered. One might guess that the sparseness of T_p would lead to considerable freedom.

V. METRIC LINEARITY

Each tangent space must be a linear space if an index formalism is to be developed. Scalar multiplication can already be defined simply by $\alpha(\beta, i, t_0) = (\alpha\beta, i, t_0)$. It turns out that it is natural to introduce the full linearity simultaneously with the metric.

Axiom 2: There exists a real-valued symmetric function g, called a metric, on $\cup_{p \in S} (T_p \otimes T_p)$ and a constant N, called the dimension of S, which satisfy

(a) For each T_p there is a finite set $\{\eta_\alpha | \alpha = 1 \to N\} \subseteq T_p$ such that for any $\xi \in T_p$ there are unique real numbers ξ^α for which

$$g(\xi, \xi') = \sum_{\alpha=1}^{N} \xi^\alpha g(\eta_\alpha, \xi')$$

is satisfied for all $\xi' \in T_p$;

(b) $g(\alpha\xi, \xi') = \alpha g(\xi, \xi')$ for all $\xi, \xi' \in T_p$, $\alpha \in R$.

In this axiom the ξ^α look like coordinates in an N dimensional linear space with basis $\{\eta_\alpha\}$. In fact, if there were a completeness condition on T_p, then the ξ^α would form a linear space. In general, however, the set of ξ^α

must be extended which leads us to the following corollary to Axiom 2.

Corollary 1: For each $p \in S$ let $\mathcal{T}_p$ be an N dimensional real linear space. There is a map $\phi_g : T_p \to \mathcal{T}_p$ such that

(a) span $\phi_g(T_p) = \mathcal{T}_p$;

(b) Whenever $\phi_g(\xi_1) = \phi_g(\xi_2)$, then $g(\xi_1, \xi) = g(\xi_2, \xi)$ for all $\xi \in T_p$;

(c) Let $\bar{g} : \phi_g(T_p) \otimes \phi_g(T_p) \to R$ be defined by $\bar{g}(\phi_g(\xi_1), \phi_g(\xi_2)) = g(\xi_1, \xi_2)$. Then $\bar{g}$ has a unique bilinear extension to $\mathcal{T}_p \otimes \mathcal{T}_p$.

Proof: Choose both an $\{\eta_\alpha\}$ set as in axiom 2 and a basis $\{V_\alpha\}$ for $\mathcal{T}_p$. Let $\phi_g(\xi) = \sum_{\alpha=1}^{N} \xi^\alpha V_\alpha$. This map satisfies the theorem.

For later reference one of the properties of ϕ_g is listed as a second corollary.

Corollary 2: $\phi_g(\alpha\xi) = \alpha\phi_g(\xi)$ for all $\alpha \in R$, $\xi \in T_p$.

For ease of notation, the extension of g to $\mathcal{T}_p$ will also be denoted g. It will be assumed as usual that g has an invertible matrix representation $g_{\alpha\beta}$ in $\mathcal{T}_p$. $\mathcal{T}_p$ will be called simply the tangent space at p, and its elements tangent vectors. Note that since $\mathcal{T}_p = \text{span}\,\phi_g(T_p)$, any linear operator is uniquely determined by its values on $\phi_g(T_p)$.

VI. DIFFERENTIABLE FUNCTIONS

A differentiable function is defined as follows:

Definition 2: Let $f : S \to R$ be a real-valued function on a space set S such that $f(i, t)$ is a differentiable function of t for all paths $i \in I$. f is then said to be differentiable on S.

It was mentioned earlier that differentiable functions would be operators on tangent vectors. Hence comes the following definition of the differential df.

Definition 3: For a differentiable function f the differential $df : \cup_{p \in S} T_p \to R$ is defined by $df(\alpha, i, t_0) = \alpha\, df/dl\,(i, t)|_{t=t_0}$

As with the metric, the differentials should be linear operators. However, it is *not* necessary for the linearity to be the same as for the metric. The linear structure for df is introduced analogously to that of g.

Axiom 3: There is a ring D of functions differentiable on S (usual addition and multiplication with inverses for nonzero functions) satisfying

(a) For each T_p there is a set $\{\bar{\eta}_\alpha | \alpha = 1 \to N\} \subseteq T_p$ (N = dimension of S) such that for any $\xi \in T_p$ there exist unique real numbers $\bar{\xi}^\alpha$ so that for any $f \in D$ $df(\xi) = \sum_{\alpha=1}^{N} \bar{\xi}^\alpha df(\bar{\eta}_\alpha)$;

(b) The constant functions are in D.

As with T_p, there is a certain sparseness allowable in D. While it will be closed under all the operations we will consider, it will still not be necessary for D to include all differentiable functions on S. An example of this type will be considered later in the paper.

The next two corollaries for df are analogous to corollaries 1 and 2 for g.

Corollary 3: For each $p \in S$ there is a map $\phi_f : T_p \to \mathcal{T}_p$ called an f map such that

(a) span $\phi_f(T_p) = \mathcal{T}_p$;

(b) Whenever $\phi_f(\xi_1) = \phi_f(\xi_2)$, $df(\xi_1) = df(\xi_2)$ for all $f \in D$;

(c) For each $p \in S$ and $f \in D$ let $\overline{df} : \phi_f(T_p) \to R$ be defined by $\overline{df}(\phi_f(\xi)) = df(\xi)$. Then $\overline{df}$ has a unique linear extension to $\mathcal{T}_p$.

Proof: Analogous to Corollary 1.

Corollary 4: $\phi_f(\alpha\xi) = \alpha\phi_f(\xi)$ for all $\alpha \in R$, $\xi \in T_p$.

The extensions of df will again be denoted df for simplicity, and the vector representation of df in $\mathcal{T}_p$ will be denoted $f_{,\alpha}$. It may be remarked here that while $f_{,\alpha}$ will share many of the usual properties of the partial derivative, it is a distinctly different operator. In particular, it will be shown later that $f_{,\alpha,\beta} - f_{,\beta,\alpha}$ need not vanish.

The situation is as in Fig. 1. There are two different types of maps on T_p embedding it in a linear space $\mathcal{T}_p$.

In Riemannian geometry the two are locked together by demanding that $\phi_f = \phi_g$. Certainly some relation is required if $g_{\alpha\beta}$ is to act on $f_{,\alpha}$ type indices, however, the relation may be much more general without sacrificing the benefits of the usual index formalism.

Axiom 4: For every point $p \in S$ there are specified an f map ϕ_f and a g map ϕ_g from T_p into $\mathcal{T}_p$.

With this axiom, the domain of df can be extended from $\cup_{p \in S} T_p$ to include $\cup_{p \in S} \mathcal{T}_p$ by using Corollary 3. There is a similar extension for the domain of g to include $\cup_{p \in S} (\mathcal{T}_p \otimes \mathcal{T}_p)$ by using Corollary 1. Strictly speaking these are new maps, but we choose to continue the former notation and just note the change of domain.

The next axiom will complete the transition from the abstract entities T_p to the familiar N-dimensional linear spaces $\mathcal{T}_p$.

VII. TENSORS

The next step is to introduce tensor fields for covariant differentiation, and an unfamiliar situation is confronted. Without a coordinate system, there are no natural, guaranteed well-behaved vector fields such as the $\partial/\partial x^j$ provided by the coordinate patches. One could resort to the $f_{,\alpha}$ vector fields and deal with neighborhoods, but this raises topological questions which are not felt to be central to the development. Consequently, to remedy this problem, a well-behaved reference basis is hypothesized.

Axiom 5: There exists a set of N(= dimension of S) functions $\{V_1, \ldots, V_N\}$ on S where $\{V_j(p)\}$ is a basis in $\mathcal{T}_p$. With respect to these bases $g_{\alpha\beta} \in D$ and for all $f \in D$ $f_{,\alpha} \in D$.

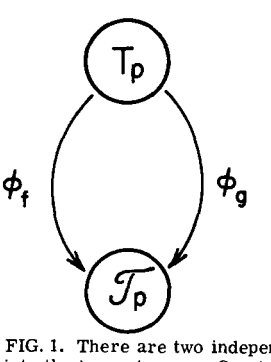


FIG. 1. There are two independent maps from the basic tangent space into the tangent space. One is generated by the metric and one by the differentiable functions.

In terms of this basis field a differentiable basis field can be defined.

Definition 4: Let $A^l_k(p)$ be an invertible $N \times N$ matrix (N = dimension of S) composed of N^2 functions in D. $\{W_k(p)\} = \{A^l_k(p)V_l(p)\}$ is then a differentiable basis field.

The stage is now set for a development closely paralleling that of normal Riemannian geometry. When the treatment is standard, it is assumed that the reader can fill in missing details, but when the treatment deviates, the presentation will be rigorous.

Tensors of rank M at p are to be linear operators on $\mathcal{T}_p \otimes \ldots \otimes \mathcal{T}_p = \mathcal{T}_p^M$, and they will have matrix representations defined by

$$R(\xi_1, \ldots, \xi_M) = R_{\alpha_1 \cdots \alpha_M} \xi_1^{\alpha_1} \cdots \xi_M^{\alpha_M}$$

Tensors of rank zero are the real numbers.

Indices will be raised and lowered using $g_{\alpha\beta}$ and its inverse $g^{\alpha\beta}$. Composition, contraction, permutation, addition, and multiplication by real numbers produces new tensors from old as usual.

A tensor field $R(p)$ will give for each p a tensor at p with some fixed rank M. With respect to a differentiable basis field, the matrix form of $R(p) = R_{\alpha_1 \cdots \alpha_M}(p)$ will always be assumed to be a differentiable function on S. $g_{\alpha\beta}, g^{\alpha\beta}$, and $f_{,\alpha}$ are already differentiable. One can check that all differentiable basis fields are equally good in determining differentiability of a tensor field.

VIII. COVARIANT DIFFERENTIATION

Covariant differentiation is critical to any extension of Riemannian geometry, and the necessary-sufficient condition for a unique differential operator will turn out to be quite simple. First, the operator is defined by listing certain ones of its usual properties.

Definition 5: d is a differential operator if it converts tensor fields of rank M $T_{\alpha_1 \cdots \alpha_M}$ into tensor fields of rank $M+1$ denoted $T_{\alpha_1 \cdots \alpha_M ; \alpha_{M+1}}$ and satisfies

(a) df is the differential of f, i.e., $f_{;\alpha} = f_{,\alpha}$

(b) $(f T_{\alpha_1 \cdots \alpha_M})_{;\alpha_{M+1}} = f_{,\alpha_{M+1}} T_{\alpha_1 \cdots \alpha_M} + f T_{\alpha_1 \cdots \alpha_M ; \alpha_{M+1}}$

(c) $(T_{\alpha_1 \cdots \alpha_M} + U_{\alpha_1 \cdots \alpha_M})_{;\alpha_{M+1}} = T_{\alpha_1 \cdots ; \alpha_{M+1}} + U_{\alpha_1 \cdots ; \alpha_{M+1}}$

(d) $g_{\alpha\beta;\gamma} = g^{\alpha\beta}{}_{;\gamma} = g^{\alpha}{}_{\beta;\gamma} = 0$

(e) $(T_{\alpha_1 \cdots \alpha_M} U^{\alpha_1 \cdots \alpha_K}{}_{\beta_1 \cdots \beta_L})_{;\alpha} = T_{\alpha_1 \cdots \alpha_M ; \alpha} U^{\alpha_1 \cdots \alpha_K}{}_{\beta_1 \cdots \beta_L} + T_{\alpha_1 \cdots \alpha_M} U^{\alpha_1 \cdots \alpha_K}{}_{\beta_1 \cdots \beta_L ; \alpha}$

(f) $f_{,\alpha;\beta} - f_{,\beta;\alpha} = 0.$

Many properties can be quickly derived from this list, i.e., $T^{\alpha}{}_{;\beta} = g^{\alpha\gamma} T_{\gamma;\beta}$ follows from (d) and (e).

To deduce the usual form of the covariant derivative using Christoffel symbols, begin by verifying

$$(fg)_{,\alpha} = f g_{,\alpha} + g f_{,\alpha} \quad \text{for } f, g \in D$$

from the definition of df.

Then for any two differentiable tensor fields ξ_α, η_β use this result plus (e) and (a) to prove that

$$g^{\alpha\beta}{}_{,\gamma} \xi_\alpha \eta_\beta + g^{\alpha\beta} \xi_{\alpha,\gamma} \eta_\beta + g^{\alpha\beta} \xi_\alpha \eta_{\beta,\gamma} = (g^{\alpha\beta} \xi_\alpha \eta_\beta)_{,\gamma} = g^{\alpha\beta} \xi_{\alpha;\gamma} \eta_\beta + g^{\alpha\beta} \xi_\alpha \eta_{\beta;\gamma}.$$

This rearranges to

$$(g^{\alpha\beta} \xi_{\alpha;\gamma} - \tfrac{1}{2} g^{\alpha\beta}{}_{,\gamma} \xi_\alpha - g^{\alpha\beta} \xi_{\alpha,\gamma}) \eta_\beta = -(g^{\alpha\beta} \eta_{\beta;\gamma} - \tfrac{1}{2} g^{\alpha\beta}{}_{,\gamma} \eta_\beta - g^{\alpha\beta} \eta_{\beta,\gamma}) \xi_\alpha.$$

Since the right side is linear in ξ_α, so is the left side. One concludes that

$$g^{\alpha\beta} \xi_{\alpha;\gamma} - \tfrac{1}{2} g^{\alpha\beta}{}_{,\gamma} \xi_\alpha - g^{\alpha\beta} \xi_{\alpha,\gamma} = C^{\beta\alpha}_{\gamma} \xi_\alpha$$

for some matrix field $C^{\beta\alpha}_{\gamma}$ independent of ξ_α.

This can be rewritten in the standard form of

$$\xi_{\alpha;\gamma} = \xi_{\alpha,\gamma} + \Gamma^{\rho}_{\alpha\gamma} \xi_\rho,$$

where $\Gamma^{\rho}_{\alpha\gamma}$ is some matrix field independent of ξ_α.

To extend this proof to tensor fields of arbitrary rank show that any tensor field can be written as a sum of terms of the form $f \xi^1_{\alpha_1} \cdots \xi^M_{\alpha_M}$, then apply (c). The usual results emerge.

One can derive the $\Gamma^{\alpha}_{\beta\gamma}$ in any of a variety of standard ways.

We begin by using $g_{\alpha\beta;\gamma} = 0$.

I. $$g_{\alpha\beta;\gamma} = g_{\alpha\beta,\gamma} + \Gamma^{\rho}_{\alpha\gamma} g_{\rho\beta} + \Gamma^{\rho}_{\beta\gamma} g_{\alpha\rho} = 0.$$

Now use the torsion-free condition (f) on a function $f \in D$:

$$f_{,\alpha;\beta} - f_{,\beta;\alpha} = (f_{,\alpha,\beta} - f_{,\beta,\alpha}) + (\Gamma^{\rho}_{\alpha\beta} - \Gamma^{\rho}_{\beta\alpha}) f_{,\rho} = 0.$$

Thus

II. $$\Gamma^{\rho}_{\alpha\beta} - \Gamma^{\rho}_{\beta\alpha} = S^{\rho}_{\alpha\beta}$$

can be determined simply by twice differentiating functions in D.

Equations I and II completely fix $\Gamma^{\rho}_{\alpha\beta}$. The easiest way to write the solution is to lower the superscripts on $\Gamma^{\rho}_{\alpha\beta}$ and $S^{\rho}_{\alpha\beta}$ to $[\rho, \alpha\beta]$ and $S_{\rho,\alpha\beta}$. Then

$$[\alpha, \beta\gamma] = -\tfrac{1}{2}(g_{\beta\alpha,\gamma} + g_{\alpha\gamma,\beta} - g_{\gamma\beta,\alpha} + S_{\beta,\alpha\gamma} + S_{\alpha,\gamma\beta} - S_{\gamma,\beta\alpha}).$$

One can now go back and calculate the covariant derivative for raised indices with the usual result. Perhaps it might be mentioned that $S^{\rho}_{\alpha\beta}$ is not a new quantity special to the generalized manifolds developed here. It appears whenever the basis vectors are freed from the $\partial/\partial x^j$ of the coordinate system and are allowed to be any complete set of independent vector fields.[3]

What has been outlined here is a uniqueness proof for the differential operator d. In fact, a complete characterization has emerged just as it does in Riemannian geometry. The sole restriction placed on the manifold resulted from the torsion-free condition (f). These results are summarized below with the generalization of a standard theorem of differential geometry.

Covariant Derivative Existence-Uniqueness Theorem

If for a differentiable basis field $\{V_\alpha\}$ there is a matrix field $S^\rho_{\alpha\beta}$ such that $f_{,\alpha,\beta} - f_{,\beta,\alpha} = S^\rho_{\alpha\beta} f_{,\rho}$ is satisfied for all $f \in D$, then and only then is there a differential operator satisfying Definition 5. Furthermore, d is unique and is defined by

$$T^{\beta_1 \cdots \beta_N}_{\alpha_1 \cdots \alpha_M ; \alpha_{M+1}} = T^{\beta_1 \cdots \beta_N}_{\alpha_1 \cdots \alpha_M, \alpha_{M+1}} + \Gamma^\rho_{\alpha_1 \alpha_{M+1}} T^{\beta_1 \cdots \beta_N}_{\rho \alpha_2 \cdots \alpha_M} + \cdots + \Gamma^\rho_{\alpha_M \alpha_{M+1}} T^{\beta_1 \cdots \beta_N}_{\alpha_1 \cdots \rho} - \Gamma^{\beta_1}_{\rho \alpha_{M+1}} T^{\rho \cdots \beta_N}_{\alpha_1 \cdots \alpha_M} + \cdots - \Gamma^{\beta_N}_{\rho \alpha_{M+1}} T^{\beta_1 \cdots \rho}_{\alpha_1 \cdots \alpha_M},$$

where $\Gamma^\rho_{\alpha\beta}$ is given by

$$\Gamma^\rho_{\alpha\beta} = -\tfrac{1}{2} g^{\rho\gamma}(g_{\alpha\gamma,\beta} + g_{\gamma\beta,\alpha} - g_{\beta\alpha,\gamma} + S_{\alpha,\gamma\beta} + S_{\gamma,\beta\alpha} - S_{\beta,\alpha\gamma})$$

and $S_{\alpha,\beta\gamma} = g_{\alpha\rho} S^\rho_{\beta\gamma}$.

IX. CURVATURE TENSOR

The curvature tensor can be found as usual by investigating $\xi^\alpha{}_{;\beta;\gamma} - \xi^\alpha{}_{;\gamma;\beta}$. Using the explicit form for the covariant derivative, one can find that

$$\xi^\alpha{}_{;\beta;\gamma} - \xi^\alpha{}_{;\gamma;\beta} = R^\alpha{}_{\rho\beta\gamma}\xi^\rho,$$

where

$$R^\alpha{}_{\beta\gamma\delta} = \Gamma^\alpha{}_{\beta\delta,\gamma} - \Gamma^\alpha{}_{\beta\gamma,\delta} + \Gamma^\alpha{}_{\rho\delta}\Gamma^\rho{}_{\beta\gamma} - \Gamma^\alpha{}_{\rho\gamma}\Gamma^\rho{}_{\beta\delta} - \Gamma^\alpha{}_{\beta\rho}S^\rho{}_{\gamma\delta}.$$

The usual tensor formulas for inverted derivative indices are also valid and can be similarly deduced.

This curvature tensor actually satisfies all the usual symmetries:

(a) $R_{\alpha\beta\gamma\delta} = -R_{\alpha\beta\delta\gamma}$,

(b) $R_{\alpha\beta\gamma\delta} = -R_{\beta\alpha\gamma\delta}$,

(c) $R_{\alpha\beta\gamma\delta} = R_{\gamma\delta\alpha\beta}$,

(d) $R_{\alpha[\beta\gamma\delta]} = 0$, [ ] denotes antisymmetrization.

(a) follows immediately from the definition. (b) follows from $h_{,\alpha;\beta} - h_{,\beta;\alpha} = 0$, where $h = g_{\tau\lambda}\xi^\tau\xi^\lambda$. (c) is implied by (a), (b), and (d). (d) then remains to be shown. A direct calculation of (d) using the explicit matrix form above gives

$$R^\alpha_{[\beta\gamma\delta]} = S^\alpha_{\beta\delta,\gamma} + S^\alpha_{\gamma\beta,\delta} + S^\alpha_{\delta\gamma,\beta} + S^\alpha_{\rho\beta}S^\rho_{\gamma\delta} + S^\alpha_{\rho\gamma}S^\rho_{\delta\beta} + S^\alpha_{\rho\delta}S^\rho_{\beta\gamma} \equiv Q^\alpha_{\beta\gamma\delta}.$$

This quantity can be shown to vanish as follows: Differentiate with respect to γ the expression

$$f_{,\alpha,\beta} - f_{,\beta,\alpha} = -S^\rho_{\alpha\beta} f_{,\rho}$$

and get

$$f_{,\alpha,\beta,\gamma} - f_{,\beta,\alpha,\gamma} = -S^\rho_{\alpha\beta,\gamma} f_{,\rho} - S^\rho_{\alpha\beta} f_{,\rho,\gamma}.$$

Antisymmetrize with respect to (α,β,γ), and wherever inverted final derivative indices appear use the S matrix. The result is (with relabelling)

$$Q^\alpha_{\beta\gamma\delta} f_{,\alpha} = 0.$$

Since f is arbitrary,

$$Q^\alpha_{\beta\gamma\delta} = 0.$$

X. BIANCHI IDENTITIES

The Bianchi identities still hold in the general case. The standard proof of Riemannian geometry works here as well.[4]

XI. EXAMPLES

Example 1: The simplest case is that of a single path i which never crosses itself:

$$G = \{(i,s) \mid s \in R\} = S.$$

The basic tangent space at the point (i,s) is

$$T_{(i,s)} = \{(\alpha, i, s) \mid \alpha \in R\}.$$

In this case $T_{(i,s)}$ is already a linear space, since all multiples of the single independent vector (l,i,s) are in $T_{(i,s)}$. To get to $\mathcal{T}_{(i,s)}$, ϕ_f and ϕ_g satisfy

$$\phi(\alpha\xi) = \alpha\phi(\xi).$$

Thus the most general ϕ_f and ϕ_g can only multiply each length by a constant.

By an appropriate choice of basis for $\mathcal{T}_{(i,s)}$, we are free to have

$$\phi_g(\alpha, i, s) = \alpha.$$

Then

$$\phi_f(\alpha, i, s) = \alpha\chi(s),$$

where $\chi(s)$ is some nonzero, differentiable function of s.

The simplest reference basis is to use 1 at every point.

Now g_{11}, the only metric component, is specified as a nonzero differentiable function of s, and the construction is finished. $S^1_{11} = 0$ trivially since $S^\rho_{\alpha\beta} = -S^\rho_{\beta\alpha}$.

D is the set of differentiable functions on R, and

$$f_{,1} = \frac{1}{\chi(s)}\frac{df}{ds}.$$

Finally,

$$\Gamma^1_{11} = -\frac{1}{2}\frac{g_{11,1}}{g_{11}} = -\frac{1}{2g_{11}\chi}\frac{dg_{11}}{ds}(s).$$

Example 2: Consider the case of R^N with N independent vector fields $\{\xi^k_{(1)}, \ldots, \xi^k_{(N)}\}$. The $\xi^k_{(\alpha)}$ can be integrated for each α to give a set of paths using

$$\frac{dx^k}{ds} = \xi^k_{(\alpha)}.$$

Here $S = R^N$. The basic tangent space T_p at each point consists of all multiples of the N vectors $\xi^k_{(\alpha)}$. The tangent space $\mathcal{T}_p$ is a copy of R^N.

To get to $\mathcal{T}_p$, again we have the restriction

$$\phi(\alpha\xi) = \alpha\phi(\xi)$$

for ϕ_f and ϕ_g. If D is to be the usual set of differentiable functions on R^N, then it is convenient to choose a basis for $\mathcal{T}_p$ so that

$$\phi^k_f(\xi_{(\alpha)}) = \xi^k_{(\alpha)}.$$

With this choice $f_{,\alpha} = \partial f/\partial x^\alpha$ and $S^\rho_{\alpha\beta} = 0$.

If $\phi_g = \phi_f$, then normal Riemannian geometry results. The most general case is to specify the vectors each $\xi_{(\alpha)}$ maps into.

$$\phi_g^k(\xi_{(\alpha)}) = A_l^k \xi_{(\alpha)}^l.$$

A_l^k is then an $N \times N$ invertible matrix (invertible because the image of the $\xi_{(\alpha)}^k$ basis must be a basis in $\mathcal{T}_p$).

Let $g_{\alpha\beta}$ be specified. $\Gamma_{\beta\gamma}^{\alpha}$ is the usual result since $S_{\alpha\beta}^{\rho} = 0$.

$$\Gamma_{\beta\gamma}^{\alpha} = -\tfrac{1}{2} g^{\alpha\rho}[g_{\rho\beta,\gamma} + g_{\rho\gamma,\beta} - g_{\beta\gamma,\rho}],$$

where from the above choice of ϕ_f, the comma denotes the usual derivative.

The expression for the curvature tensor is the accustomed one as well.

At this point one might wonder if anything new has been added since the usual expressions are appearing. The difference is in the metric. The metric defined by $ds^2 = \bar{g}_{\alpha\beta} dx^\alpha dx^\beta$ is not $g_{\alpha\beta}$ but is $\bar{g}_{\alpha\beta} = g_{\mu\nu} A_\alpha^\mu A_\beta^\mu$. This follows from

$$g(\xi_{(\alpha)}, \xi_{(\beta)}) = g_{\mu\nu}\phi_g^\mu(\xi_{(\alpha)})\phi_g^\nu(\xi_{(\beta)}) = (g_{\mu\nu} A_k^\mu A_l^\nu)\xi_{(\alpha)}^k \xi_{(\beta)}^l.$$

Thus the metric $\bar{g}_{\alpha\beta}$ which defines proper time in terms of the dx^α is not the metric $g_{\alpha\beta}$ which appears in parallel propagation or in the curvature tensor.

The last example will show that the Brans–Dicke scalar–tensor field equations are a special case of this geometry where $A_\alpha^\mu = \pm \phi^{1/2}\delta_\alpha^\mu$.

Example 3: This example deals with R^N again but with $M > N$ vector fields $\{\xi_{(1)}^\alpha, \ldots, \xi_{(M)}^\alpha\}$. ($N$ of these are to be independent.) The results are analogous to Example 2 with the difference that ϕ_g can now be a more complex map, a nonlinear map. In other words, there may not be a matrix $\bar{g}_{\mu\nu}$ so that

$$g(\xi_{(\alpha)}, \xi_{(\beta)}) = \bar{g}_{\mu\nu}\xi_{(\alpha)}^\mu \xi_{(\beta)}^\nu.$$

It is this generalization to increasingly complex ϕ_g maps which allows many more fields than the Brans–Dicke scalar field to be introduced. Their scalar–tensor theory is but the simplest step beyond Riemannian geometry.

Example 4: This example has $M < N$ and illustrates a very different kind of freedom by constructing a two-dimensional manifold which *fills* a three-dimensional space. One way to do this is to start in R^3 and specify two independent, infinitely differentiable vector fields $\xi_{(1)}^k, \xi_{(2)}^k$. (Latin indices will run from 1 to 3, Greek from 1 to 2.) At each point $p \in R^3$, T_p is the multiple of these two vectors, and $\mathcal{T}_p$ is a copy of R^2. These vector fields can be integrated to give a set of parametrized paths filling R^3.

Now begin at point p_0 and ask what other points can be reached by traveling from p_0 along the path net. Consider infinitesimal translation. Certainly one can use two infinitesimal translations (one on a $\xi_{(1)}$ path and one on a $\xi_{(2)}$ path) to move infinitesimally in any direction spanned by $\xi_{(1)}^k$ and $\xi_{(2)}^k$ at p_0. Now consider the following displacements to second order. Move Δs along a $\xi_{(1)}$ path, Δs along a $\xi_{(2)}$ path, $-\Delta s$ along a $\xi_{(1)}$ path, $-\Delta s$ along a $\xi_{(2)}$ path. To first order in Δs one has returned to the starting point. However, to second order the displacement from p_0 is

$$\Delta x^k = \left(\frac{\partial \xi_{(2)}^k}{\partial x^l}\xi_{(1)}^l - \frac{\partial \xi_{(1)}^k}{\partial x^l}\xi_{(2)}^l\right)\Delta s^2 = I^k \Delta s^2.$$

If one were dealing with the paths of a two-dimensional hypersurface in R^3, then Δx^k would be in the tangent space spanned by $\xi_{(1)}^k$ and $\xi_{(2)}^k$, since one would stay in the hypersurface. In general, however, I^k has a component normal to the tangent space which means that infinitesimal translations in *all* directions are possible from p_0. This guarantees that the paths *cannot* be separated into two-dimensional hypersurfaces and that the two-dimensional manifold fills a three-dimensional volume.

To continue the construction of the manifold let $\phi_g^\beta(\xi_{(\alpha)}) = A_\alpha^\beta(p)$ and let $\phi_f^\beta(\xi_{(\alpha)}) = \delta_\alpha^\beta$. Any differentiable function f on R^3 produces a differential operator df where

$$df(\eta) = f_{,\alpha}\eta^\alpha = \left(\frac{\partial f}{\partial x^k}\xi_{(\alpha)}^k\right)\eta^\alpha.$$

Thus $f_{,\alpha} = \partial f/\partial x^k\, \xi_{(\alpha)}^k$ and $S_{\alpha\beta}^{\rho}$ can be calculated.

$$f_{,\alpha,\beta} - f_{,\beta,\alpha} = \frac{\partial f}{\partial x^k}\left(\frac{\partial \xi_{(\alpha)}^k}{\partial x^l}\xi_{(\beta)}^l - \frac{\partial \xi_{(\beta)}^k}{\partial x^l}\right)\xi_{(\alpha)}^l = \left(\frac{\partial f}{\partial x^k}\xi_{(\rho)}^k\right)\left[\xi_m^{(\rho)}\left(\frac{\partial \xi_{(\alpha)}^m}{\partial x^l}\xi_{(\beta)}^l - \frac{\partial \xi_{(\beta)}^k}{\partial x^l}\xi_{(\alpha)}^l\right)\right],$$

where the inverse matrix $\xi_m^{(\rho)}$ has been used. The inverse matrix is not a full inverse since that is impossible for a 3×2 matrix. Thus $\partial f/\partial x^k$ must be restricted. For simplicity it is assumed that $\hat{x}^3$ does not lie in the span of $\xi_{(1)}^k$ and $\xi_{(2)}^k$ and that D is all infinitely differentiable functions on R^3 satisfying $\partial f/\partial x^3 \equiv 0$, i.e., $f = f(x_1, x_2)$.

In this case a unique inverse to $\xi_{(\rho)}^k$ exists satisfying

$$\xi_{(\rho)}^k \xi_l^{(\rho)} = \delta_l^k \quad \text{for } k = 1, 2\ l = 1, 2, 3.$$

Thus $\xi_{(\rho)}^k \xi_l^{(\rho)}$ is the identity on the $\hat{x}^1, \hat{x}^2$ tangent space. Now one has

$$S_{\alpha\beta}^{\rho} = \xi_m^{(\rho)}\left(\frac{\partial \xi_{(\alpha)}^m}{\partial x^l}\xi_{(\beta)}^l - \frac{\partial \xi_{(\beta)}^m}{\partial x^l}\xi_{(\alpha)}^l\right).$$

Closure of D under differentiation gives the sole restriction on the basis field $\{\xi_{(\alpha)}^k\}$ of

$$\frac{\partial \xi_{(\alpha)}^k}{\partial x^3} = 0 \quad \text{for } k = 1, 2.$$

Now introduce an invertible 2×2 $g_{\alpha\beta}$ matrix composed of functions in D and the construction is finished. $\Gamma_{\beta\gamma}^{\alpha}$ can be calculated as usual as can the curvature tensor.

It might be pointed out that if $A_\beta^\alpha \in D$, then all tensor fields in the formalism are functions of x^1 and x^2 alone. The tensor structure then reduces to that of a two-dimensional manifold of the type in Example 2. However, A_β^α, which is a differentiable tensor field in its own right, is under no such restrictions.

Example 5: (Brans–Dicke scalar–tensor theory): A special case of Example 2 leads to familiar results when field equations are sought. Let $A_\beta^\alpha = \pm \phi^{+1/2}\delta_\beta^\alpha$ so that $g_{\alpha\beta} = \bar{g}_{\alpha\beta}/\phi$.

If one calculates the scalar curvature R, one finds

$$R = \phi\bar{R} - 3\bar{\Box}\phi + \frac{9}{2}\frac{\phi_{,\alpha}\phi_{,\beta}}{\phi}\bar{g}^{\alpha\beta},$$

where $\bar{R}$ and $\bar{\Box} = {}_{,\alpha}{}^{;\alpha}$ are calculated as if $\bar{g}_{\alpha\beta}$ were the metric. Now make R a scalar density by multiplying by $\sqrt{-\bar{g}}$. (If one multiplied by $\sqrt{-g}$, the variational equations for ϕ and $\bar{g}_{\alpha\beta}$ are dependent which gives an underdetermined system.) The result is a Lagrangian density equal to that of Brans–Dicke with their constant $w = -9/2$ plus one term $-3\bar{\Box}\phi\sqrt{-\bar{g}}$.

$$\mathfrak{L} = \int\left(\phi\bar{R} + \frac{9}{2}\frac{\phi_{,\alpha}\phi_{,\beta}\bar{g}^{\alpha\beta}}{\phi} - 3\bar{\Box}\phi\right)\sqrt{-\bar{g}}\,d^4x.$$

However, $\bar{\Box}\phi\sqrt{-\bar{g}} = (\phi\cdot{}^{\alpha}\sqrt{-\bar{g}})_{,\alpha}$ and is a simple divergence which drops out of the variational equations. Vary ϕ and $\bar{g}^{\alpha\beta}$ to get the usual Brans–Dicke field equations.

The advantage of this approach is that the scalar field ϕ is not introduced *ad hoc*, but is a natural part of the generalized geometry.

A similar result was discovered by Sen and Dunn[5] who worked with a Lyra manifold where the displacement vector from x^μ to $x^\mu + dx^\mu$ was $\xi^\mu = x^0 dx^\mu$. This is reminiscent of the map $\xi^\alpha = \phi^{1/2}\bar{\xi}^\alpha$ from ϕ_f to ϕ_g. Again the problem with the Lyra manifold is that it does not seem to be a natural extension of Riemannian geometry.

As with Sen and Dunn, while the field equations for $\bar{g}_{\alpha\beta}$ are the same as those of Brans–Dicke, the geodesics are different. This comes in our case from $g_{\alpha\beta}$ (not $\bar{g}_{\alpha\beta}$) giving the Christoffel symbols. Thus a full comparison of this theory with that of Brans–Dicke would be rather more involved.

Perhaps it might be pointed out again that the scalar transformation used in this section is but the simplest possibility beyond a usual Riemannian manifold. Many other fields will appear as the transform is made more general. Another point is that the Brans–Dicke field equations for any value of w can be gotten by multiplying the above Lagrangian density by an appropriate power of ϕ.

XII. CONCLUSIONS

The coordinates have been removed from Riemannian geometry and replaced by a path net. In addition, the metric and derivative operators were allowed to induce different linearities on the tangent space. By doing this, considerable generalization has emerged without losing any of the basic Riemannian formalism such as Christoffel symbols, curvature tensor, or Bianchi identities.

The derivation of the generalized manifold was presented in a hopefully natural seeming way by building in those properties considered useful and seeing what structure emerged. The Brans–Dicke field equations were shown to be the next simplest step beyond Riemannian manifolds.

In this process some potentially interesting questions were bypassed. Topological questions were not considered in the derivation for two reasons. First, unlike with the usual coordinate-patch presentation of Riemannian geometry, topology seemed to be of secondary importance and was not necessary to introduce the theory. In addition, a consideration of topology looked like a complex enough subject for a paper of its own. Since it is possible to construct a two-dimensional manifold filling a three-dimensional space (or four, or five, $\cdots$), the topological characterization would seem likely to be peculiar.

The exact role of the coordinate system has also been sidestepped. Although coordinates are not presented as fundamental, in all the examples thus far studied, a coordinate system has been used to provide a matrix for the reference-path net. Until an example without coordinates is constructed, or proved not to exist, the role of the coordinates will remain open.

[1]Several other generalizations of Riemannian geometry can be found in H. Weyl, *Space, Time, and Matter* (Dover, New York, 1922), M. Tonnelat, *Einstein's Theory of Unified Fields* (Gordon and Breach, New York, 1966), G. Lyra, Math. Z. **54**, 52 (1951).

[2]C. Brans and R. Dicke, Phys. Rev. **124**, 925 (1961).

[3]As an exmaple of this S matrix in Reimannian geomerry (although with different notation) see H. Flanders, *Differential Forms* (Academic, New York, 1963), p. 129.

[4]R. Adler, M. Bazin, and M. Schiffer, *Introduction to General Relativity* (McGraw Hill, New York, 1965), p. 142.

[5]D. Sen and K. Dunn, J. Math. Phys. **12**, 569 (1971).

Scattering theory in a model of quantum fields. I

Sergio Albeverio*†

Department of Physics, Joseph Henry Laboratories, Princeton University, Princeton, New Jersey 08540
(Received 2 May 1972)

We study a model of quantum field theory with "Yukawa-like" interaction $\lambda \int \Phi_b^+(\mathbf{x})\Phi_b^-(\mathbf{x})\Phi_a(\mathbf{x})d\mathbf{x}$ between nucleons (b) and mesons (a). It is a version of Nelson's model with relativistic kinematics and has been renormalized by J. P. Eckmann. The infinite mass renormalization is a power series in λ^2, chosen in such a way as to confer on the renormalized Hamiltonian $\hat{H}$ the correct relativistic single particle spectrum. Physical one nucleon states are given by a modified Friedrichs one-particle expansion constructed by Eckmann. The Heisenberg picture's creation–annihilation operator for dressed nucleons and mesons are studied in detail, as a preparation for the construction of the correspondent asymptotic fields, carried through, in this paper, for the mesons fields in general and for the nucleon fields on particular states (the general case is treated in the second paper of this series). Analytic properties of the interacting fields in λ are proved and commutation relations of the asymptotic fields are established. Moreover, strong asymptotic states are constructed as well as isometric wave operators. Finally some reduction formulas for the meson–nucleon scattering are derived.

1. INTRODUCTION

The mathematical formulation of the physical description of scattering processes of multiparticle quantum mechanical systems can be separated into two steps. The first consists in solving the existence problem for the basic asymptotic quantities of scattering theory (e.g. wave operators, asymptotic fields, S matrix). The second step includes the proof of some important general properties (like unitarity and asymptotic completeness). In nonrelativistic problems with 2-body potentials of short range, step one is rather well under control. In step 2, however, difficulties are met for systems with $N \geq 3$ particles (see, e.g., Ref. 1 and the references given there).

The situation is of course even much more complicated in quantum field theory. Here even the first step gives troubles.[2] In general we may roughly say that once a solution of the "one-body problem"[3] is given, then the asymptotic quantities can be constructed, provided the interaction has some "locality properties." This is the case, e.g., in the framework of a local, relativistic Wightman's theory[4] where under correspondent assumptions on the spectrum of the energy–momentum operator asymptotic states can be constructed as strong limits, for $t \to \pm\infty$, by applying suitable time-dependent "one-particle excitation operators," the existence of which is postulated, to the vacuum.[5]

The convergence of the "interacting" fields to asymptotic free fields[6,7] has also been proved in this framework on a set of states.[8,9] However, despite recent impressive progress, especially in two-dimensional models,[10] the justification in a constructive way of the different assumptions underlying such theories is still not completed. Somehow, between the two mentioned "extreme" cases of models for nonrelativistic N particle systems and local relativistic interacting quantum fields, there is a set of "intermediate models" in which existence of asymptotic fields and/or states has been studied from a mathematical point of view. These models retain only one or another subset of the features and difficulties of local relativistic quantum field theory with interesting scattering. Among these let us mention for our purposes two main classes of models with infinitely many degrees of freedom[11]:

(a) models which have a spatially cut-off local relativistic invariant interaction (these models have vacuum problems, but no one-particle problems)[12];

(b) "persistent models" which have one-particle problems and number of particle divergences, but no vacuum problems.[13]

"Vacuum" (resp. "one particle problems") refer to the difficulties connected with the construction of the physical vacuum (resp. the physical one particle) states.

The models which we shall study in this paper are of the class (b). A first example of a model of this class is Nelson's model of nonrelativistic nucleons coupled linearly with a boson field.

It describes scalar relativistic mesons and nonrelativistic nucleons with a Yukawa-like interaction. The formal Hamiltonian of the model is $H_f = H_0 + \lambda V_f$, where H_0 is the second quantization operator for the total kinetic energy in the Fock space of the nucleons and mesons and, using $\Psi_b(x)$ for the second quantized nucleon field, $\Phi_a(x)$ for the meson field, $V_f = \int \Psi_b^*(x)\Phi_a(x)\Psi_b(x)dx$, λ being a real number (the "coupling constant").

This model, which has a logarithmic divergence (in 4-dimensional space–time), has been renomalized by Nelson,[14] so to obtain a dynamics given by a lower bounded, self-adjoint Hamiltonian (in any sector with finitely many nucleons, the number of nucleons being conserved). Cannon[15] has shown that, at least for small total momenta and small coupling constant, the vacuum is an isolated point of the spectrum of the Hamiltonian and an isolated physical one-nucleon state (eigenstate of the Hamiltonian in the center of mass system) can be constructed by analytic perturbation theory. However, the energy of this state and its detailed dependence on mass and momentum are left open by this investigation.[16]

Høegh-Krohn has shown the existence of asymptotic (bare ≡ dressed) meson fields.[17] We shall see from the related Eckmann's models we shall study in detail that the asymptotic limit of nucleon fields is much more complicated.[18] These models are versions of Nelson's model with relativistic kinematics, describing scalar mesons and nucleons in (2, 3 or) 4 space–time dimensions with a "polarization free" Yukawa interaction.

The formal Hamiltonian is

$$H_f = H_0 + \lambda V_f, \tag{1.1}$$

where H_0 is the total kinetic energy in the Fock space of nucleons and mesons and

$$V_f = \int \Phi_b^+(x)\Phi_b^-(x)\Phi_a(x)dx, \tag{1.2}$$

$\Phi_b^+(x)$ resp. $\Phi_b^-(x)$ being the creation resp. annihilation part of the nucleon field and $\Phi_a(x)$ being the meson field, λ a real number (the coupling constant). In 4 (and 3) space–time dimensions renormalization is a necessity already in order to get a well-defined Hamiltonian.[19] The general procedure is to introduce an ultraviolet cut off $\sigma < \infty$ in the interaction and an additive mass renormalization $\mathfrak{M}_\sigma$ in such a way that $H_0 + \lambda V_\sigma + \mathfrak{M}_\sigma(\lambda)$ is self-adjoint, bounded from below on states with finitely many nucleons and gives in a suitable limit for $\sigma \to \infty$ a Hamiltonian with the same properties. It is easily seen that this leaves a great deal of ambiguity in the definition of $\mathfrak{M}_\sigma$.

In particular, different choices of $\mathfrak{M}_\sigma$ give in general different dependence of the one nucleon energy $E_\lambda(q)$ on the momentum q. In this paper we shall concentrate on a choice $\hat{M}_\sigma$ of $\mathfrak{M}_\sigma$ which makes $E_\lambda(q) = (m_b^2 + q^2)^{1/2}$ and thus yield the correct relativistic one particle spectrum. The construction of $\hat{M}_\sigma$ has been carried out with high skill by Eckmann,[20] who was able to show that, at least for bounded nucleon momenta and suitably small coupling constant, $\hat{M}_\sigma$ can be given in the form of a power series in λ^2, with operator coefficients. The term proportional to λ^2 diverges as $\sigma \to \infty$ (infinite mass renormalization), whereas the terms of higher order remain bounded. The Hamiltonian $\hat{H}_\sigma = H_0 + \lambda V_\sigma + \hat{M}_\sigma$ is self-adjoint and bounded from below. The one nucleon (improper) eigenstates of $\hat{H}_\sigma$ are obtained by applying a suitable dressing transformation (given by a Friedrichs-type expansion) to the bare one nucleon states. Formally, such an eigenstate (which we call an "unnormalized dressed or physical one nucleon state") is given by an expression of the following kind:

$$\hat{b}^*(q)\Omega_0 = b^*(q)\Omega_0 + \sum_{l=1}^{\infty} b_l^*(q)\Omega_0 \tag{1.3}$$

and satisfies, by construction, formally $\hat{H}_\sigma \hat{b}^*(q)\Omega_0 = \omega(q)\hat{b}^*(q)\Omega_0$, where Ω_0 is the vacuum, $b^*(q)$ the creation operator for a bare nucleon of momentum q and $b_l^*(q)$ have the operator form $b^*(a^*)^l$ (with certain kernels) [a^* being the creation operator for mesons and $(a^*)^l$ standing for the product of l such operators]. To get normalized one nucleon physical states (of Fock space norm 1, when smeared with suitable functions) one has still to perform an amplitude renormalization (see Ref. 20) and, for the corresponding situation in Lee models, K. Hepp, Ref. 21, Ch. III). In this paper we first define (Sec. 1), without actually using any specific form of $\hat{M}_\sigma$, the Heisenberg picture adjusted[22] meson creation and annihilation operators $\hat{a}^\#_{\sigma,t}(h) \equiv e^{it\hat{H}_\sigma}e^{-itH_0}a^\#(h)e^{itH_0} \times e^{-it\hat{H}_\sigma}$, where h is a square integrable function and $a^\#(h)$ stands for the creation operator $a^*(h)$ or the annihilation operator $a(h)$. We then show that the strong limits as $t \to \pm\infty$ of these operators exist and have the usual commutation relations of asymptotic fields,[23] between themselves and with the Hamiltonian. These results hold for different choices of $\hat{M}_\sigma$, in particular for the one of Ref. 20b and the one we shall study further in the following sections. In Sec. 3 we introduce, for a specific choice of $\hat{M}_\sigma$, the correspondent creation and annihilation operators $\hat{b}^\#$ for "dressed nucleons",[24] which act on the vacuum in the way given formally by (1.3). Moreover, a field strength renormalization has to be introduced in addition to the dressing. Due to the highly unbounded character of the dressed nucleon fields [because of the presence of the "meson cloud" acting as $\sum b_l^*(q)\Omega_0$ on Ω_0] some care must be taken to show that the fields are actually defined in every subspace with finitely many nucleons. We study the properties of these dressed fields as well as of the meson fields in the Heisenberg picture given by $\hat{H}_\sigma$. We prove then the strong asymptotic convergence of time dependent states constructed with these fields, containing finitely many dressed nucleons and infinitely (or finitely) many mesons. We also show strong asymptotic convergence of certain operators to partial isometric "wave operators," which have all the properties one expects for wave operators from our knowledge of N-body systems.[25]

The results are extended to the case without any cutoff.

The study of the asymptotic convergence of the adjusted Heisenberg picture dressed nucleon fields is more complicated, and will be tackled in part II of this work.[26] This study will be based on the domain properties and on estimates on the dressed nucleon fields we establish in 3A of this paper. So it may be useful for a better understanding of this section to give, already at this point, a motivation for the procedure which shall be used in part II for constructing asymptotic physical nucleon fields.

The general procedure[27] in the method of asymptotic fields[23] for a system described by an Hamiltonian H which is, in a suitable sense, the sum of H_0 and a perturbation W, is to get sufficient decay in t of a norm of the form $I \equiv \|K_t e^{-itH}\Psi\|$, where Ψ is in a suitable dense set of vectors in the relevant Fock space and K_t is a certain commutator [essentially of W and $C^\#(e^{\mp itH}h)$, where $C^\#(e^{\mp itH}h)$ are the adjusted creation–annihilation operators for the field under study, h a test function and Ω the relevant one particle energy].

In the case of space cutoff models, $C^\#(\cdot)$ can be taken to be a "bare field" and therefore K_t is equal to a bounded operator after multiplication by suitable inverse powers of the number operator (and/or H_0). Then, using so called "higher order estimates"[28] and decay properties of smooth solutions of the Klein–Gordon equation,[29] the wanted estimate on I follows.

However, in the model we study in this series of papers, without space cutoff, $C^\#(\cdot)$ cannot be taken to be a bare field.[30] In part II we generalize the method of asymptotic fields by taking $C^\#(\cdot)$ as a "dressed field", in our case a dressed nucleon field. Then K_t is a sum of infinitely many terms $K_t^{(l)}$, $l = 0, 1, \cdots$ of the above form, but control on the sum can still be obtained on a suitable dense set of states. For this procedure the results of 3A of the present paper are essential.

In the subsequent Sec. 3B we introduce the S matrix and establish some relations between asymptotic states and fields.

Finally in 3C we derive some reduction formulas for the meson–nucleon scattering.

In part II we shall continue the investigation by constructing, as mentioned above, strong asymptotic physical nucleon fields, studying their properties, in particular establishing their commutation relations with the asymptotic meson fields constructed in this paper. Also other results on the S matrix and the meson–nucleon scattering will be given.

We would like to mention that the methods used in this series of papers can be applied, with due modifications, to a class of other models (including Nelson's model, Eckmann's and Nelson's models with Fermi statistics for nucleons and/or some internal group).

Finally, we would like to point out that J. Fröhlich,[16]

working independently with other methods, has obtained a lot of interesting results in a class of models with persistent vacuum, which are partly complementary to ours.

Notations

As mentioned in the Introduction, the model describes two types of particles, which are called "nucleons" or "b particles" and "mesons" or "a particles". A label a resp. b attached to a quantity will always denote that this quantity is used for mesons resp. nucleons. We shall assume that the number of space dimensions in which the particles move and interact is three. All results hold, however, with suitable adaptations, also for $s = 2, 1$.[31] Later on we shall need the distinction between the so called "bare" and "physical" (or "dressed") nucleons. We make the general convention of understanding simply under nucleons always the bare ones, unless the qualification "physical" is added.

Let $\mathfrak{F}_a, \mathfrak{F}_b$ denote the Fock spaces for the mesons alone resp. the nucleons alone. We shall assume that both the mesons and nucleons in the model have Bose–Einstein statistics (although most of the results can be extended to the case where the nucleons are fermions and/or some internal group is present). Thus, in our present case, $\mathfrak{F}_a, \mathfrak{F}_b$ are both isomorphic ($\cong$) to the well-known symmetric Fock space for bosons

$$\mathfrak{F} = \bigoplus_{\nu=0}^{\infty} \mathfrak{F}^{(\nu)},$$

$\oplus$ denoting direct sum and $\mathfrak{F}^{(\nu)}$ being the Hilbert space $L_2^{(s)}(\mathbb{R}^{3\cdot\nu})$ of symmeric square integrable functions over the $3\cdot\nu$-dimensional Euclidean space $\mathbb{R}^{3\cdot\nu}$. Due to the fact that the interaction we are going to study conserves the number of nucleons,[32] we shall mostly make all considerations in a fixed subspace ("sector")

$$\mathfrak{K}^{(n)} = \mathfrak{F}_b^{(n)} \otimes \mathfrak{F}_a, \quad n = 0, 1, 2, \cdots \tag{1.4}$$

of $\mathfrak{K}$, consisting of all vectors with the fixed number n of nucleons ($\mathfrak{F}_b^{(n)} \cong L_2^s(\mathbb{R}^{3n})$). One has $\mathfrak{K}^{(n)} = \oplus_{m=0}^{\infty} \mathfrak{K}^{(n,m)}$, where $\mathfrak{K}^{(n,m)}$ is the "subspace with n nucleons and m mesons" (subspace of $L_2(\mathbb{R}^{3(m+n)})$ consisting of functions which are separately symmetric in all nucleons and mesons arguments). We shall always consider L_2-functions in momentum space. We write $(\cdot)$ resp. $\|\ \|$ for the scalar product resp. the norm in $\mathfrak{K}$ as well as in the L_2-spaces we might consider, whenever no confusion is to be feared. We shall call N_a resp. N_b the number operators for mesons resp. nucleons and set $N = N_a + N_b$ for the total number operator. The creation–annihilation operator for mesons ($a^{\#}(h)$) and nucleons ($b^{\#}(h)$) are defined in the usual way, for all $h \in L_2(\mathbb{R}^3)$. (If C, C^* is any pair of operators we shall use $C^{\#}$ as a short way to include both in a statement: Thus $C^{\#}$ stands for "C^* or C".) The normalization is such that $[a(h), a^*(g)] = (h, g)$, where we used the notation $[A, B] \equiv AB - BA$ for any two operators A, B. We shall call $a^{\#}(k), b^{\#}(k), (k \in \mathbb{R}^3)$ the formal operators associated in the usual way with the operators $a^{\#}(h), b^{\#}(h)$.

The free Hamiltonian H_0, describing "bare nucleons," each with mass $m_b > 0$, and mesons, each with mass $m_a > 0$, is given by

$$\begin{aligned} H_0 &= H_0^{(a)} + H_0^{(b)}, \\ H_0^{(a)} &= \int \mu(k)\, a^*(k)\, a(k)\, dk, \\ H_0^{(b)} &= \int \omega(k)\, b^*(k)\, b(k)\, dk, \end{aligned} \tag{1.5}$$

where $\mu(k) \equiv (m_a^2 + |k|^2)^{1/2}, \omega(k) \equiv (m_b^2 + |k|^2)^{1/2}$ are the kinetic energy of a meson resp. nucleon of momentum k ($|k|$ being the Euclidean norm). H_0 is a positive, self-adjoint operator on its natural domain of definition $D_1 \equiv D(H_0)$ in $\mathfrak{K}$.

Let us now describe the interaction. Formally it is given by (1.2), where

$$\Phi_a^{\pm}(x) = (2\pi)^{-3} \int dk \mu(k)^{-1/2} e^{\pm ikx} a^{\#}(k), \tag{1.6}$$

$$\Phi_b^{\pm}(x) = (2\pi)^{-3} \int dk \omega(k)^{-1/2} e^{\pm ikx} b^{\#}(k), \tag{1.7}$$

$+$ going with $a^*, b^*, -$ with a, b and $\Phi_a(x) = \Phi_a^+(x) + \Phi_a^-(x)$, $\Phi_b(x) = \Phi_b^+(x) + \Phi_b^-(x)$ being the free meson fields resp. free "bare" nucleon fields at point x. If we rewrite (1.2) explicitly in terms of the operators $a^{\#}(k), b^{\#}(k)$ in momentum space, we obtain

$$V = V^c + V^a, \tag{1.8}$$

with

$$V^c = \int \omega(k_1)^{-1/2} \omega(k_2)^{-1/2} \mu(k_3)^{-1/2} \delta(k_1 - k_2 + k_3) \times b^*(k_1) b(k_2) a^*(k_3)\, dk_1 dk_2 dk_3 \tag{1.9}$$

$$V^a = \int \omega(k_1)^{-1/2} \omega(k_2)^{-1/2} \mu(k_3)^{-1/2} \delta(k_1 - k_2 - k_3) \times b^*(k_1) b(k_2) a(k_3)\, dk_1 dk_2 dk_3. \tag{1.10}$$

It is easy to see that any nonzero vector in $\mathfrak{K}^{(n)}, n > 0$ with finite number of mesons, is not in the domain of V. One will get control on the operator after introduction of a momentum cutoff, as we are going to see in Secs. 2 and 3. First, however, let us make a general remark on the terminology. In the following we shall have to ask the question whether a given Hamiltonian H' (obtained by addition to $H_0 + \lambda V$ of some suitable mass renormalization) has the relativistic spectrum of nucleons respectively mesons energies. Let us explain this concept using a decomposition of the Fock space and all operators described, e.g., in Refs. 15, 16, and 20. Since H_0 and V conserve the total momentum P, also H' will of course conserve P. Hence H' will be reduced in the decomposition of the Fock space $\mathfrak{K}$ and of each sector $\mathfrak{K}^{(n)}$ into a direct integral of Hilbert spaces $\mathfrak{K}_P$ resp. $\mathfrak{K}_P^{(n)}$ corresponding to given P. Call H'_P the restriction of H' to $\mathfrak{K}_P$. We say that H' has the "(physical) relativistic spectrum of nucleon energies" when there exists a vector $\varphi \in \mathfrak{K}_P^{(1)}$ such that $H'_P\varphi = \omega(P)\varphi$. φ is then called a physical (or dressed) one nucleon state [with energy $\omega(P)$] and we speak of φ as being a state of a physical nucleon. We say also that φ is an (improper) eigenstate of H' to the correct relativistic energy for the physical nucleon. Correspondingly, H' is said to have the (physical) relativistic spectrum of meson energies when there exists a vector $\Psi \in \mathfrak{K}_P^{(0)}$ such that $H'_P\Psi = \mu(P)\Psi$, Ψ being then called a physical one meson state [with energy $\mu(P)$]. Ψ is also called an (improper) eigenstate of H' to the correct energy of the mesons. If one replaces everywhere in above definition H'_P by $H_0|_P$ one has the so called bare one nucleon resp. one meson states. Thus, in short, bare particles are associated with the free Hamiltonian H_0, physical particles with the total Hamiltonian H'. Due to the properties of the interaction all H' we shall consider will always have the property $H' = H_0$ on $\mathfrak{K}^{(0)}$, which implies that all physical meson states coincide with bare meson states (any Ψ as above also satisfies $H_0|_P\Psi = \mu(P)\Psi$ and belongs to the bare meson–no nucleons subspace $\mathfrak{K}^{(0,1)}$. Thus, in the models we shall consider, there is no difference between bare and physical mesons. This is not so for the nucleons;

physical nucleons states will not belong to $\mathcal{H}^{(1,0)}$. Of course a state Φ in $\mathcal{H}^{(1)}$ will be called physical (or dressed) when its reduction to $\mathcal{H}_P^{(1)}$ is a physical one nucleon state.

Linear combinations, closed linear hulls, and tensor products of such states will have the adjective "dressed" attached to them, as well as any other associated quantity, like, e.g., the fields we shall construct in Sec. 3.[33] An operator H' will be said to have the relativistic one particle spectrum if it has the relativistic spectrum of both nucleons and mesons energies.

2. SOME SIMPLE ESTIMATES. THE MESON FIELDS AND THE CORRESPONDENT ASYMPTOTIC FIELDS

Consider the interaction V as given by (1.8), (1.9), (1.10). Let $\sigma \geq 0$ and let us call $V_\sigma, V_\sigma^c, V_\sigma^a$ the operators we obtain from these expressions for V, V^c, V^a replacing the functions $\omega(k)^{-1/2}$, $\mu(k)^{-1/2}$ in the integrands by the cutoff functions $\omega_\sigma(k)^{-1/2} \equiv \chi_\sigma(k)\omega(k)^{-1/2}$, $\mu_\sigma(k)^{-1/2} \equiv \chi_\sigma(k) \times \mu(k)^{-1/2}$, where $\chi_\sigma(\cdot)$ is a cutoff function over $\mathbb{R}^3$ defined by: $\chi_\sigma(k) = \chi(|k| - \sigma)$, where $\chi(\cdot)$ is a infinitely differentiable (C^∞) function on the real line $\mathbb{R}$ which is 0 in the interval $[0, \infty)$, 1 in the interval $(-\infty, -1)$, and satisfies, for all $u \in \mathbb{R}$, $0 \leq \chi(u) \leq 1$.

We call σ the ultraviolet cutoff. We assume always $\sigma < \infty$, unless stated otherwise.

Let $\hat{M}_\sigma$ be an operator in $\mathcal{H}$ of the form

$$\hat{M}_\sigma = \int m_\sigma(\lambda, k)\, b^*(k)\, b(k)\, dk, \tag{2.1}$$

where $m_\sigma(\lambda, k)$ is, for any $\sigma < \infty$, a bounded continuous function of k. Then $\hat{M}_\sigma$ is a bounded operator in each sector $\mathcal{H}^{(n)}$. Later on, a particular form of m_σ will be assumed.

Set, for $\sigma < \infty$,

$$\hat{H}_\sigma = H_0 + \lambda V_\sigma + \hat{M}_\sigma, \tag{2.2}$$

where λ is the coupling constant.

In this and the subsequent sections we shall often restrict operators A, defined in $\mathcal{H}$, to the subspace $\mathcal{H}^{(n)}$ and denote their restrictions by $A|_n$ (or sometimes $A|_{\mathcal{H}^{(n)}}$). Since the number of nucleons is conserved by the interaction, such restrictions are quite natural.

Lemma 2.1: For any $0 \leq \sigma < \infty$ and all real λ, the following estimate holds for any $\Psi \in D(H_0) \cap \mathcal{H}^{(n)}$:

$$\|\lambda V_\sigma \Psi\| \leq 2n|\lambda| L_\sigma \|(N_a + 1)^{1/2}\Psi\| \leq \epsilon \|H_0 \Psi\| + C(\epsilon)\|\Psi\|, \tag{2.3}$$

for any $\epsilon > 0$, where

$$L_\sigma \equiv \left(\sup_p \int \omega_\sigma^{-1}(p+q)\, \omega_\sigma^{-1}(p)\, \mu_\sigma^{-1}(q)\, dq \right)^{1/2} < \infty \tag{2.4}$$

and $C(\epsilon) \equiv m_a \epsilon + (m_a \epsilon)^{-1}(2n|\lambda| L_\sigma)^2 < \infty$.

Moreover, setting $V'_\sigma \equiv \lambda V_\sigma + \hat{M}_\sigma$,

$$\|V'_\sigma \Psi\| \leq 2n|\lambda| L_\sigma \|(N_a + 1)^{1/2}\Psi\| + \|\hat{M}_\sigma|_n\| \, \|\Psi\| \leq \epsilon \|H_0 \Psi\| + C'(\epsilon)\|\Psi\|, \tag{2.5}$$

for any $\epsilon > 0$, where $\|\hat{M}_\sigma|_n\|$ is the norm of the bounded operator $\hat{M}_\sigma|_n$ and one has $\|\hat{M}_\sigma|_n\| = n \cdot L_\sigma^1$, $L_\sigma^1 < \infty$ being a constant independent of n. Furthermore,

$$C'(\epsilon) \equiv C(\epsilon) + \|\hat{M}_\sigma|_n\| = C(\epsilon) + n \cdot L_\sigma^1.$$

In particular, $\lambda V_\sigma|_n$ are relatively small perturbations of $H_0|_n$ with relative bound less than 1 (in the terminology of Ref. 34, p. 190).

The following "first order estimates" will also be used:

$$\|V'_\sigma \Psi\| \leq K_1(\sigma)\|(\hat{H}_\sigma + K_2(\sigma))\Psi\| \tag{2.6}$$

and

$$\|H_0 \Psi\| \leq K_3(\sigma)\|(\hat{H}_\sigma + K_4(\sigma))\Psi\|, \tag{2.7}$$

for any $\Psi \in D(H_0) \cap \mathcal{H}^{(n)}$ and some finite constants $K_1(\sigma)$, $K_2(\sigma)$, $K_3(\sigma)$, $K_4(\sigma)$ (independent of Ψ).

Proof: This is essentially Lemma 3 in Ref. 20b and easily proved using the definition of V_σ, the trivial estimate $\|N_a^{1/2}\varphi\| \leq (1/m_a)\|H_0^{(a)1/2}\varphi\|$ for any $\varphi \in D(H_0^{(a)1/2})$, together with the simple fact that if A is a self-adjoint operator, then $\|A\varphi\| \leq \epsilon\|A^2\varphi\| + (1/\epsilon)\|\varphi\|$ for any $\epsilon > 0$, $\varphi \in D(A^2)$. ∎

From this lemma one has (Ref. 20b):

Theorem 2.1: In each sector $\mathcal{H}^{(n)}$, for any $0 \leq \sigma < \infty$ and all λ, the cutoff Hamiltonian $\hat{H}_\sigma|_n$ defined by

$$\hat{H}_\sigma|_n = H_0|_n + \lambda V_\sigma|_n + \hat{M}_\sigma|_n = H_0|_n + V'_\sigma|_n, \tag{2.8}$$

on $D(H_0) \cap \mathcal{H}^{(n)}$, is self-adjoint and lower bounded.

The perturbations $\lambda V_\sigma|_n$, $\hat{M}_\sigma|_n$, $V'_\sigma|_n$ are symmetric operators defined on domains which contain $D(H_0|_n)$.

Define $\hat{H}_\sigma$ in $\mathcal{H}$ as the operator the restriction of which to $\mathcal{H}^{(n)}$ is precisely $\hat{H}_\sigma|_n$. $\hat{H}_\sigma$ is self-adjoint on the domain $D(H_0)$ and is bounded below on every $\mathcal{H}^{(n)}$ (and thus in all subspaces with a finite number of nucleons). ∎

In Ref. 20, Eckmann has shown that there exist choices of the function m_σ such that suitable limits of $\hat{H}_\sigma|_n$ for $\sigma \to \infty$ exist and define self-adjoint operators $\hat{H}_\infty|_n$, which are lower bounded in each sector $\mathcal{H}^{(n)}$; e.g., for the choice $m_\sigma(\lambda, k) = \lambda^2 \omega_\sigma(k)^{-1} \int \mu_\sigma(q)^{-1} \omega_\sigma(q-k)^{-1}[\mu(q) + \omega(q-k)]^{-1} dq$, the norm limit $R_\infty(z)$ of $(z - \hat{H}_\sigma|_n)^{-1}$ exists for all λ, with any choice of z such that $\mathrm{Re}\, z < -\delta(\lambda, n)$, where $\delta(\lambda, n) > 0$ is independent of σ.

Moreover, $z - R_\infty(z)^{-1} \equiv \hat{H}_\infty|_n$ is then independent of z and is self-adjoint on the range of $R_\infty(z)$ and lower bounded in $\mathcal{H}^{(n)}$. In a simple way one can then define $\hat{H}_\infty$ as the self-adjoint operator which has the restriction $\hat{H}_\infty|_n$ in $\mathcal{H}^{(n)}$.

By a well-known theorem of Trotter one has that e^{itH_σ} converges strongly in $\mathcal{H}$ to $e^{it\hat{H}_\infty}$, as $\sigma \to \infty$.

This choice of m_σ gives what we may call the "simplified model." This model has, however, the disadvantage of not having the correct relativistic spectrum of the physical one nucleon energies.[35] In Sec. 3 we shall study the "full model" in which another choice of m_σ is made, in order to get the correct physical spectrum.

For the rest of this section we shall simply assume that a choice C of m_σ is made in such a way that $e^{it\hat{H}_\sigma|_n}$ converges strongly as $\sigma \to \infty$ to $e^{it\hat{H}_\infty|_n}$, where $\hat{H}_\infty|_n$ is self-adjoint. Let $\mathcal{D} \equiv \mathcal{D}(\mathbb{R}^3) \equiv C_0^\infty(\mathbb{R}^3)$ be Schwartz's test function space of infinitely continuously differentiable functions of compact support in momentum space $\mathbb{R}^3$. Define $\mathcal{D}_{(\sigma)} \equiv L_2(\mathbb{R}^3)$ for $\sigma < \infty$, $\mathcal{D}_{(\infty)} \equiv \mathcal{D}$. The usual notation $(T)^*$ for the adjoint of an operator T will be used.

We can now state the following two theorems:

Theorem 2.2: Let $n \geq 0$ be a fixed integer and let $H_0, \hat{H}_\sigma, a^\#(h)$ be the restrictions to $\mathcal{K}^{(n)}$ of the kinetic energy resp. the Hamiltonian resp. the meson creation-annihilation operators defined in Sec. 1 and (2.8). Assume the above choice C for m_σ is made.[36] Set, for any $h \in L_2(\mathbb{R}^3), \sigma \leq \infty$:

$$\hat{a}^\#_{\sigma,t}(h) \equiv e^{it\hat{H}_\sigma} e^{-itH_0} a^\#(h) e^{itH_0} e^{-it\hat{H}_\sigma}. \tag{2.9}$$

We call $\hat{a}^\#_{\sigma,t}(h)$ the "adjusted" meson creation and annihilation operators in the Heisenberg picture given by $\hat{H}_\sigma$.

Then for all $h \in \mathfrak{D}_{(\sigma)}$, the following holds:

1. $\hat{a}^\#_{\sigma,t}(h)$ has the same domain of definition $D(a^\#(h))$ as $a^\#(h)$ and for any $\Psi \in D(H_0^{1/2})$ one has $\hat{a}^\#_{\sigma,t}(h)\Psi - a^\#(h)\Psi = \mathfrak{A}^\#_\sigma(h,t)\Psi$, where $\mathfrak{A}^\#_\sigma(h,t) \equiv i\lambda \int_0^t e^{iu\hat{H}} \sigma \mathfrak{N}^\#_\sigma(h_{\pm u}) e^{-iu\hat{H}_\sigma} du$ and $a^\#(h_{\pm u}) \equiv e^{-iuH_0} a^\#(h) e^{iuH_0}, h_{\pm u} \equiv e^{\mp i\mu t}h, h_+$ going with a^*, h_- with a. Moreover, $\mathfrak{A}^\#_\infty(h_{\pm u}) = \text{n-}\lim_{\sigma\to\infty} \mathfrak{A}^\#_\sigma(h_{\pm u})$, $\mathfrak{N}^\#_\sigma(h_{\pm u})$ being the bounded extension of $[V_\sigma, a^\#(h_{\pm u})]$ to all $\mathcal{K}^{(n)}$. $\mathfrak{A}^\#_\sigma(h;t)$ for all $\sigma \leq \infty$ is a bounded operator, uniformly bounded in t, which converges in norm as $t \to \pm\infty$ to the bounded operators $\mathfrak{A}^\#_\sigma(h, \pm\infty) = i\lambda \int_0^{\pm\infty} e^{iu\hat{H}_\sigma} \mathfrak{N}^\#_\sigma(h_{\pm u}) e^{-iu\hat{H}_\sigma} du$ (all integrals should be understood in the strong sense).

2. As $t \to \pm\infty$, $\hat{a}^\#_{\sigma,t}(e)$ converges strongly for all $\sigma \leq \infty$, on $D(H_0^{1/2})$, to the limits $\hat{a}^\#_{\sigma,\pm}(h)$ and one has $\hat{a}^\#_{\sigma,\pm}(h) = a^\#(h) + \mathfrak{A}^\#_\sigma(h;\pm\infty)$. The linear operators $\hat{a}^\#_{\sigma,\pm}(h)$ map $D(\hat{a}^\#_{\sigma,\pm}(h)) \cap \mathcal{K}^{(n,m)}$ into $\mathcal{K}^{(n,m\pm 1)}, m+1$ going with a^*, $m-1$ with a.

Moreover, for all t, $\Psi \in D(a^*(h)) \cap \mathcal{K}^{(0)}$:

$$\hat{a}^*_{\sigma,\pm}(h)\Psi = \text{s-}\lim_{s\to\pm\infty} \hat{a}^*_{\sigma,s}(h)\Psi = \hat{a}^*_{\sigma,t}(h)\Psi = a^*(h)\Psi \tag{2.10}$$

and

$$\hat{a}_{\sigma,\pm}(h)\Omega_0 = \text{s-}\lim_{s\to\pm\infty} \hat{a}_{\sigma,s}(h)\Omega_0 = \hat{a}_{\sigma,t}(h)\Omega_0 = a(h)\Omega_0 = 0. \tag{2.11}$$

Equations (2.10) and (2.11) hold (also in the case $\sigma = \infty$) for all $h \in L_2(\mathbb{R}^3)$.

Theorem 2.3:

1. The asymptotic meson creation and annihilation operator $\hat{a}^\#_{\sigma,\pm}(h)$ of Theorem 2.2 can be extended, under the same assumption $h \in \mathfrak{D}_{(\sigma)}, \sigma \leq \infty$, to have the same domain of definition as the original meson creation and annihilation operators $a^\#(h)$. We denote the extended, closed operators by the same symbols $\hat{a}^\#_{\sigma,\pm}(h)$. Then one has that $\hat{a}^*_{\sigma,\pm}(h)$ is the adjoint of $\hat{a}_{\sigma,\pm}(\bar{h})$ [denoting by $\bar{h}$ the function "complex conjugate" to h: $\bar{h}(q) \equiv \overline{h(q)}$] and one has

$$D(\hat{a}^*_{\sigma,\pm}(h)) = D(\hat{a}_{\sigma,\pm}(h)) = D(a^*(h)) = D(a(h)).$$

2. $e^{it\hat{H}_\sigma}$ and $\hat{a}^\#_{\sigma,\pm}(h)$ satisfy the same commutation relations as do e^{itH_0} and $a^\#(h)$:

$$\begin{aligned} e^{-it\hat{H}_\sigma} \hat{a}^*_{\sigma,\pm}(h) e^{it\hat{H}_\sigma} &= \hat{a}^*_{\sigma,\pm}(e^{-it\mu}h), \\ e^{-it\hat{H}_\sigma} \hat{a}_{\sigma,\pm}(h) e^{it\hat{H}_\sigma} &= \hat{a}_{\sigma,\pm}(e^{it\mu}h). \end{aligned} \tag{2.12}$$

For $\sigma < \infty$ we have the stronger statements

$$\begin{aligned} [\hat{H}_\sigma, \hat{a}^*_{\sigma,\pm}(h)] &= \hat{a}^*_{\sigma,\pm}(\mu h), \\ [\hat{H}_\sigma, \hat{a}_{\sigma,\pm}(h)] &= -\hat{a}_{\sigma,\pm}(\mu h) \quad \text{on } D(H_0), \end{aligned} \tag{2.12'}$$

for all h such that $\mu h \in L_2(\mathbb{R}^3)$.

3a. The asymptotic annihilation and creation operators $\hat{a}^\#_{\sigma,\pm}(h)$ satisfy on a dense subset[37] $\mathring{\Delta}'^{(n)}_\sigma$ of $\mathcal{K}^{(n)}$ the same commutation relations as the original annihilation and creation operators:

$$\begin{aligned} &[\hat{a}_{\sigma,\pm}(\bar{h}), \hat{a}^*_{\sigma,\pm}(g)] = (h,g), \\ &[\hat{a}_{\sigma,\pm}(h), \hat{a}_{\sigma,\pm}(g)] = [\hat{a}^*_{\sigma,\pm}(h), \hat{a}^*_{\sigma,\pm}(g)] = 0, \end{aligned} \tag{2.13}$$

for any $g, h \in \mathfrak{D}_{(\sigma)}$. For $\sigma < \infty$, $\mathring{\Delta}'^{(n)}_\sigma \supset D(H_0)$.

3b. For any $g, h \in \mathfrak{D}_{(\sigma)}$ and $\Phi, \Psi \in D(a^\#(\bar{h})) \cap D(a^\#(g))$ the asymptotic annihilation and creation operators $\hat{a}^\#_{\sigma,\pm}(h)$ satisfy the weak commutation relations

$$((\hat{a}^\#_{\sigma,\pm}(\bar{h}))^*\Psi, \hat{a}^\#_{\sigma,\pm}(g)\Phi) - ((\hat{a}^\#_{\sigma,\pm}(g))^*\Psi, \hat{a}^\#_{\sigma,\pm}(\bar{h})\Phi) = 0, \tag{2.14}$$

where the four symbols $a^\#$ stand either all for a^* or all for a, and

$$(\hat{a}^*_{\sigma,\pm}(h)\Psi, \hat{a}^*_{\sigma,\pm}(g)\Phi) - (\hat{a}_{\sigma,\pm}(\bar{g})\Psi, \hat{a}_{\sigma,\pm}(\bar{h})\Phi) = (h,g). \tag{2.15}$$

Remark 2.1: In the case $\sigma < \infty$ one has the following bound, uniform in t and norm continuous in the test functions $h \in L_2(\mathbb{R}^3)$:

$$\|\hat{a}^\#_{\sigma,t}(h)\Psi\| \leq \text{const}\,\|h\|\,\|(\hat{H}_\sigma + \text{const})\Psi\|, \tag{2.16}$$

for all $\Psi \in D(H_0)$, which is the basis for the stronger results for $\sigma < \infty$.

Proof of the *Theorems 2.2, 2.3:* The proof is very similar to the one of the analogous propositions for the meson fields in Nelson's model given by R. Høegh-Krohn (Ref. 17).

First the relation

$$\frac{d}{dt}(\Phi, \hat{a}^\#_{\sigma,t}(h)\Psi) = i(\Phi, e^{it\hat{H}_\sigma}[V'_\sigma, a^\#(h_{\pm t})]e^{-it\hat{H}_\sigma}\Psi) \tag{2.17}$$

is easily proven, for any $\sigma < \infty$, $h \in \mathfrak{D}_{(\sigma)}$, $\Phi, \Psi \in D_1 \equiv D(H_0)$, using

$$\|a^\#(g)\Psi\| \leq \text{const}\,\|g\|\,\|(\hat{H}_\sigma + \text{const})\Psi\|. \tag{2.18}$$

Then the following lemma is established:

Lemma 2.2: Set for any $\sigma < \infty$ and $h \in \mathfrak{D}_{(\sigma)}$

$$\mathfrak{N}'^\#_\sigma(h_{\pm t}) \equiv [V'_\sigma, a^\#(h_{\pm t})].$$

Then:

(1) For all $\sigma < \infty$, $\mathfrak{N}'^\#_\sigma(h_{\pm t})$ has a dense domain of definition containing, e.g., D_1. On D_1 one has

$$\mathfrak{N}'^\#_\sigma(h_{\pm t}) = \lambda[V_\sigma, a^\#(h_{\pm t})].$$

$\mathfrak{N}'^\#_\sigma(h_{\pm t})$ is bounded on this domain (as an operator from $D_1 \cap \mathcal{K}^{(n)}$ into $\mathcal{K}^{(n)}$) and can therefore be uniquely extended in a continuous way to a bounded operator $\mathfrak{N}^\#_\sigma(h_{\pm t})$ from $\mathcal{K}^{(n)}$ into $\mathcal{K}^{(n)}$.

(2) For given h and t, $\mathfrak{N}^\#_\sigma(h_{\pm t})$ is norm convergent on $\mathcal{K}^{(n)}$, as $\sigma \to +\infty$.

Define $\mathfrak{N}^\#_\infty(h_{\pm t}) \equiv \text{n-}\lim \mathfrak{N}^\#_\sigma(h_{\pm t})$ for $\sigma \to \infty$.

(3) For all $\sigma \leq \infty$, $\mathfrak{N}^\#_\sigma(h_{\pm t})$ is strongly continuous in t.

(4) $\|\mathfrak{N}^\#_\sigma(h_{\pm t})\| \leq C^{(1)}(1 + |t|)^{-3/2}$ for all t, where the constant $C^{(1)}$ is independent of t.

Proof: (1) For $\sigma < \infty$, both $V'_\sigma a^\#(h_{\pm t})$ and $a^\#(h_{\pm t})V'_\sigma$ are defined on D_1, as seen by the estimates of Lemma

2.1 and (2.16). Since $\hat{M}_\sigma$ commutes with $a^\#(h_{\pm t})$ on D_1, we have $\mathfrak{A}'^\#_\sigma(h_{\pm t}) = \lambda[V_\sigma, a^\#(h_{\pm t})]$ on D_1.

From this we can easily compute, for $\Psi \in D_1$,

$$(\mathfrak{A}^\#_\sigma(h_{\pm t})\Psi)^{(n,m)}(\underline{q}) = \sum_{j=1}^{n} \chi_\sigma(q_j)\omega(q_j)^{-1/2}\Phi_j^{(n,m)}(\underline{q}), \qquad (2.19)$$

where q stands for the set $(q_1, \ldots, q_n)$ of nucleon variables (we do not write the meson variables since they play just the role of parameters in the entire proof) and

$$\Phi_j^{(n,m)}(\underline{q}) \equiv \int dk g_{\sigma,\pm t}(k - q_j)\Omega_j^{(n,m)}(k, q_1, \ldots, \hat{q}_j, \ldots, q_n), \qquad (2.20)$$

$$g_{\sigma,\pm t}(k') \equiv \chi_\sigma(k')\mu(k')^{-1/2}h(\pm k')e^{\mp i\mu(k')t}, \qquad (2.21)$$

$$\Omega_j^{(n,m)}(k, q_1, \ldots, \hat{q}_j, \ldots, q_n) \equiv \chi_\sigma(k)\omega(k)^{-1/2}\Psi^{(n,m)}(k, q_1, \ldots, \hat{q}_j, \ldots, q_n), \qquad (2.22)$$

where the hat $\hat{}$ on a momentum variable means omission of this variable. We use the fact that $\Phi_j^{(n,m)}$ is a convolution and take its Fourier transform with respect to the relevant variable q_j.

We have the estimate

$$\|\Phi_j^{(n,m)}\| \leq C^{(2)}\|\tilde{g}_{\sigma,\pm t}\|_\infty \|\Psi^{(n,m)}\|, \qquad (2.23)$$

where $\|\tilde{g}_{\sigma,\pm t}\|_\infty \equiv \operatorname{esssup}_x |\tilde{g}_{\sigma,\pm t}(x)|$ is the L_∞-norm of

$$\tilde{g}_{\sigma,\pm t}(x) \equiv (2\pi)^{-3/2} \int dq \chi_\sigma(q)\mu(q)^{-1/2}h(\pm q)e^{iqx \mp i\mu(q)t}.$$

From this (1), (2), (3) follow using the compactness of the support of h and Lebesgue's dominated convergence theorem. (4) follows from (2.23) and the well-known estimates on smooth solutions of the Klein–Gordon equation.[29] ∎

From this lemma point 1 of Theorem 2.2 follows immediately. Moreover, point 2 follows also, for $\sigma < \infty$ and all $h \in \mathfrak{D}$. The result extends then to all $h \in L_2(\mathbb{R}^3)$ by using (2.16).

The rest of the proof of Theorems 2.2, 2.3 follows then as in Ref. 17, using the norm convergence of $\mathfrak{A}^\#_\sigma(h_{\pm t})$ as $\sigma \to +\infty$. Point 3 of Theorem 2.3 is discussed, for the case $\sigma = \infty$, in the Appendix. ∎

Remark 2.2: The above theorems give in particular the asymptotic meson fields and their commutation relations for the "simplified model" of Eckmann discussed in Ref. 20b, for both the ultraviolet cut-off case and the case without any cut-offs. It also gives the same results for the choice of mass renormalization of next section ("full model").

Remark 2.3: Since $\hat{H}_\sigma = H_0$ on $\mathfrak{K}^{(0)}$, for any $\sigma < \infty$, we have $e^{it\hat{H}_\sigma} = e^{itH_0}$ on $\mathfrak{K}^{(0)}$ and from the choice of $\hat{M}_\sigma$, s-$\lim_{\sigma\to\infty} e^{it\hat{H}_\sigma} = e^{it\hat{H}_\infty} = e^{it\hat{H}_0}$ on $\mathfrak{K}^{(0)}$. Hence $\hat{H}_\infty = H_0$ in $\mathfrak{K}^{(0)}$. In particular, $\Omega_0 \in D(\hat{H}_\infty)$ and $\hat{H}_\infty\Omega_0 = 0$. Hence the vacuum is an eigenvector of $\hat{H}_\sigma$ for $\sigma \leq \infty$ to the eigenvalue 0 ("persistence of the vacuum").

In $\mathfrak{K}^{(0)}$ there are no other eigenvalues of $\hat{H}_\sigma$; the spectrum of $\hat{H}_\sigma$, in $\mathfrak{K}^{(0)}$, being equal to that of $H_0^{(a)}$, consists of the simple isolated eigenvalue 0 and the pure absolutely continuous part $[m_a, \infty)$. In all $\mathfrak{K}^{(n)}, n > 0$ the spectrum of $\hat{H}_\sigma$ is purely continuous (due to translation invariance).

3. THE SCATTERING QUANTITIES FOR THE HAMILTONIAN WITH THE CORRECT RELATIVISTIC MESON AND NUCLEON SPECTRUM

A. The asymptotic states and fields

The Hamiltonian and the dressed fields

We shall study at the same time the model with an ultraviolet cut-off $0 \leq \sigma < \infty$ and the model without any cut-off ($\sigma = \infty$). Whereas the complete mass renormalization will be infinite for $\sigma = \infty$, finite for $\sigma < \infty$, the amplitude renormalization will be finite in all cases. For all cases $0 \leq \sigma \leq \infty$ the renormalizations are necessary, as we shall see, in order to obtain the correct asymptotic states and fields, with the correct normalization.[38] The "one-body problem" of constructing one-nucleon dressed states has been solved by Eckmann (Ref. 20a).

We shall present its solution in a way convenient for later use. We have to introduce some notations and operators, first as formal quantities, whose domain of definition will be specified later. It will be convenient to use ordered pairs of labels (σ, σ'), where $0 \leq \sigma \leq \infty$ is an "ultraviolet cut-off in all variables" and can take any value in $[0, \infty]$ for $\sigma < \infty$ and $\sigma' = R$ is an arbitrarily given, nonnegative, finite constant. The label σ' will correspond to an ultraviolet cut-off entering *only* in the variables appearing in some nucleon *annihilation* operator.

Let

$$W^{(i)} \equiv \int \chi_\sigma(q; p_1, \ldots, p_i)w^{(i)}(q; p_1, \ldots, p_i)b^*(q - \Sigma) \times \left(\prod_{j=1}^{i} a^*(p_j)\right) b(q)dqdp_1 \cdots dp_i, \qquad (3.1)$$

for $i = 1, 2, \ldots$, where the "numerical kernel" $w^{(i)}$ is some function of $q, p_1, \ldots, p_i$ and $\chi_\sigma(q, p_1 \cdots p_i) \equiv \prod_{k=1}^{i} \chi_\sigma(p_k)\,\chi_\sigma(q)\chi_{\sigma'}(q)$, the cut-off function $\chi_\sigma(\cdot)$ being as in Sec. 2. Σ stands for $\sum_{j=1}^{i} p_j$.

Then the Friedrichs operation $\Gamma W^{(i)}$ is defined by

$$\Gamma W^{(i)} = \int \chi_\sigma(q; p_1, \ldots, p_i)w'^{(i)}(q; p_1, \ldots, p_i) \times b^*(q - \Sigma) \prod_{j=1}^{i} a^*(p_j)b(q)dqdp_1 \ldots dp_i, \qquad (3.2)$$

where $w'^{(i)}(q; p_1, \ldots, p_i) \equiv w^{(i)}(q; p_1, \ldots, p_i)(\omega(q - \Sigma) + \sum_{j=1}^{i} \mu(p_j) - \omega(q))^{-1}$. Formally, therefore, Γ acts on $W^{(i)}$ as the inverse operation to ad H_0. It is convenient to describe the action of ad H_0 resp. Γ on $W^{(i)}$ by saying that the kernel of ad $H_0(W^{(i)})$ resp. $\Gamma W^{(i)}$ is the same as the one of $W^{(i)}$ multiplied by the "difference of the energies created by the creation operators minus the sum of the energies destroyed by the annihilation operators contained in $W^{(i)}$ resp. the inverse of that sum." Define $W^{(0)}$ by

$$W^{(0)} = \int \chi_\sigma(q)\chi_{\sigma'}(q)w^{(0)}(q)b^*(q)b(q)dq. \qquad (3.3)$$

The $W^{(i)}$ are particular "Wick monomials" we shall consider often below.[39] We are going to introduce special symbols for those "contractions" between Wick monomials which we shall meet in the course of our considerations. So we define

$$W^{(0)} \underset{1}{\smile} W^{(1)} \equiv \int \chi_\sigma(q')\chi_{\sigma'}(q')\delta(q' - q + \Sigma)\chi_\sigma(q; p_1, \ldots, p_i) \times w^{(i)}(q; p_1, \ldots, p_i)b^*(q')\left(\prod_{j=1}^{i} a^*(p_j)\right) b(q)dq'dqdp_1 \cdots dp_i.$$

$W^{(0)} \underset{1}{\smile} W^{(i)}$ is the expression obtained from the product of Wick monomials $W^{(0)}W^{(i)}$ by "contracting the b annihilators in $W^{(0)}$ with the b^*-creator in $W^{(i)}$" [in general, contraction of an annihilation operator $c(p) = a(p)$ or $b(p)$ with its correspondent creation operator $c^*(p') = a^*(p')$ or $b^*(p')$ means replacing $c(p)c^*(p')$ by $\delta(p - p')$.]

In a similar way we define

$$V^a_\sigma \underset{1}{\smile} W^{(i)} = \int \omega_\sigma(k_1)^{-1/2}\omega_\sigma(k_2)^{-1/2}\mu_\sigma(k_1 - k_2)^{-1/2} \times \delta(k_2 - q + \Sigma)\chi_\sigma(q;p_1,\dots,p_i) \times w^{(i)}(q;p_1,\dots,p_i)b^*(k_1)\Big(\prod_{j=1}^{i} a^*(p_j)\Big)b(q)a(k_1 - k_2) \times dk_1dk_2dqdp_1\cdots dp_i.$$

Thus $V^a_\sigma \underset{1}{\smile} W^{(i)}$ is the term in the Wick ordered expansion of $V^a_\sigma \cdot W^{(i)}$ (given by the Wick theorem) which is obtained by contracting in the product $V^a_\sigma \cdot W^{(i)}$ the b annihilator in V^a_σ with the b^* creator in $W^{(i)}$ (and then Wick ordering).

We define also

$$V^a_\sigma \underset{2}{\smile} W^{(i)} = \sum_{l=1}^{i} \int \omega_\sigma(k_1)^{-1/2}\omega_\sigma(k_2)^{-1/2}\mu_\sigma(k_1 - k_2)^{-1/2}\delta(k_3 - p_l) \times \delta(k_2 - q + \Sigma)\chi_\sigma(q;p_1,\dots,p_i)w^{(i)}(q;p_1,\dots,p_i) \times b^*(k_1)\Big(\prod_{j=1, j\neq l}^{i} a^*(p_j)\Big)b(q)dk_1dk_2dqdp_1\dots dp_i.$$

This is the sum of all terms in the Wick ordered expansion of $V^a_\sigma \cdot W^{(i)}$ involving two contractions, i.e., all terms obtained by contracting in $V^a_\sigma \cdot W^{(i)}$ the a annihilator in V^a_σ with any of the i creators a^* in $W^{(i)}$ and then contracting the b annihilator in V^a_σ with the b^* creator in $W^{(i)}$.

We are going to define quantities which depend on the ordered pairs of labels (σ, σ'). We make everywhere the convention of omitting to write the label σ' if $\sigma < \infty$, $\sigma' = \infty$. Define, for integer, nonnegative ν, i:

$$\begin{aligned} &S_{\sigma 1,i}(\sigma') \equiv 0, \quad i = 0,2,3,4,\cdots, \\ &S_{\sigma \nu,i}(\sigma') \equiv 0, i > \nu \text{ or } i = 0, \\ &S_{\sigma 1,1}(\sigma') \equiv -\lambda V^c_\sigma(\sigma'), \end{aligned} \tag{3.4}$$

where

$$V^c_\sigma(\sigma') \equiv \int \omega(k_1)^{-1/2}\omega(k_2)^{-1/2}\mu(k_3)^{-1/2}\delta(k_1 - k_2 + k_3) \times \chi_\sigma(k_1)\chi_\sigma(k_2)\chi_\sigma(k_3)\chi_{\sigma'}(k_2)b^*(k_1)b(k_2)a^*(k_3)dk_1dk_2dk_3.$$

Then, recursively, we define

$$S_{\sigma\nu,i}(\sigma') \equiv -\lambda V^c_\sigma \underset{1}{\smile} \Gamma S_{\sigma\nu-1,i-1}(\sigma') - \lambda V^a_\sigma \underset{2}{\smile} \Gamma S_{\sigma\nu-1,i+1}(\sigma') - \sum_{k=1}^{(\nu-i)/2} M_{\sigma 2k}(\sigma') \underset{1}{\smile} \Gamma S_{\sigma\nu-2k,i}(\sigma'),$$

where

$$M_{\sigma 2k}(\sigma') \equiv -\lambda V^a_\sigma \underset{2}{\smile} \Gamma S_{\sigma 2k-1,1}(\sigma'), \quad k = 1,2,\cdots. \tag{3.5}$$

Remark that in above quantities the first index (ν resp. $2k$) denotes the order in the coupling constant λ (order ν means term proportional to λ^ν), the second index (i) the the number of meson creators.

Note also that all quantities $S_{\sigma\nu,i}(\sigma'), M_{\sigma 2k}(\sigma')$ are of the form (3.1) resp. (3.3).

Set

$$\hat{M}'_\sigma(\sigma') \equiv \sum_{k=2}^{\infty} M_{\sigma 2k}(\sigma'), \quad \hat{M}_\sigma(\sigma') \equiv \lambda^2 M_{\sigma 2}(\sigma') + \hat{M}'_\sigma(\sigma'),$$
$$\hat{H}_\sigma(\sigma') \equiv H_0 + \lambda V_\sigma + \hat{M}_\sigma(\sigma'). \tag{3.6}$$

Note that $\hat{M}'_\sigma(\sigma')$ is a sum of terms of even order $4, 6, 8, \cdots$ in λ.

Set furthermore

$$Q_\sigma(\sigma') \equiv \sum_{\nu=0}^{\infty}\sum_{i=0}^{\nu} S_{\sigma\nu,i}(\sigma') \tag{3.7}$$

and

$$T_\sigma(\sigma') \equiv :\exp(\Gamma(Q_\sigma(\sigma'))): \equiv \sum_{j=0}^{\infty} (j!)^{-1}:(\Gamma(Q_\sigma(\sigma')))^j:. \tag{3.8}$$

For the normalization of physical nucleon states we shall need an operator $A_\sigma(\sigma')$ such that $\|T_\sigma(\sigma')A_\sigma(\sigma')b^*(g)\Omega_0\| = 1$ for any $g \in L_2(\mathbf{R}^3)$, $\|g\| = 1$, Ω_0 being the vacuum in $\mathcal{K}$. It is easily seen formally using the definition of $T_\sigma(\sigma')$ that $A_\sigma(\sigma')$ can be chosen as

$$A_\sigma(\sigma') \equiv (:\exp\{[\Gamma(Q_\sigma(\sigma'))^*\Gamma(Q_\sigma(\sigma'))]_M\}:)^{-1/2},$$

where $[\ ]_M$ means the sum over all "mass graphs", i.e., all completely contracted graphs of $\Gamma(Q_\sigma(\sigma'))^*\Gamma(Q_\sigma(\sigma'))$. $A_\sigma(\sigma')$ acts on each $\mathcal{K}^{(n,m)}$ as the identity on the meson variables and as a multiplication operator in the nucleon variables. More precisely, for $\Psi^{(n,m)} \in \mathcal{K}^{(n,m)}$:

$$(A_\sigma(\sigma')\Psi)^{(n,m)}(q_1\dots q_n;p_1\dots p_m) = \prod_{i=1}^{n} \nu_\sigma(\sigma') \times (q_i)^{-1/2}\Psi^{(n,m)}(q_1\dots q_n;p_1\dots p_m), \tag{3.9}$$

where $\nu_\sigma(\sigma')(q)$ [which depends, as $A_\sigma(\sigma')$ and all the quantities (3.4)–(3.8), on λ] is the (for small λ convergent) sum of the infinitely many kernels (of even degrees in λ) one obtains by computing[40]

$$(T_\sigma(\sigma')b^*(h)\Omega_0, T_\sigma(\sigma')b^*(g)\Omega_0) = \int \nu_\sigma(\sigma')(q)\bar{h}(q)g(q)dq. \tag{3.10}$$

We shall now give domains of definition for the different operators we introduced (3.4)–(3.9). For this we need the following definition:

Let, for $\alpha > 0$, $\underline{D}^{(n);\alpha}$ be the dense subset of $\mathcal{K}^{(n)}$ defined by

$$\underline{D}^{(n);\alpha} \equiv \{\Psi^{(n)} = (\Psi^{(n,m)})_{m=0,1,2,\dots} \mid \lim_{m\to\infty} e^{\alpha m}\|\Psi^{(n,m)}\| = 0\}$$

and let $D^{(n);\alpha}$ be the Banach space obtained from $\underline{D}^{(n);\alpha}$ by taking the closure with respect to the norm $\|\Phi^{(n)}\|_\alpha \equiv \sup_m e^{\alpha m}\|\Phi^{(n,m)}\|$.

Remark 3.1: The Banach spaces $D^{(n);\alpha}$ have been introduced by Lanford in his discussion of a cut-off Yukawa theory.[41] They form a scale in the sense $D^{(n);\alpha} \supset D^{(n);\beta}$ for $0 < \alpha < \beta$. An immediate consequence of their definition is also that

$$\|\Psi\| \leq (1 - e^{-2\beta})^{-1/2}\|\Psi\|_\beta \quad \text{for any } \Psi \in D^{(n);\beta}, \quad 0 < \beta,$$

$\|\ \|$ being as usual the Fock space norm of Ψ (as element of $\mathcal{K}^{(n)}$). Another easy consequence is that N^k is for any $k > 0$ a bounded operator from $D^{(n);\alpha}$ to $D^{(n);\alpha^1}$ for any $0 < \alpha^1 < \alpha$. Moreover, $\|\Psi\|_\alpha \leq \|e^{\alpha N^a}\Psi\|$ for any $\alpha > 0$. The following subsets, which we shall use later, are contained in $D^{(n);\alpha}$ for all $\alpha > 0$. They are dense in $D^{(n);\alpha}$ in the $\|\ \|_\alpha$-topology for all $\alpha > 0$ and they are also dense in $\mathcal{K}^{(n)}$ in the Fock space topology:

(a) $\overset{0}{\mathcal{H}}{}^{(n)} \equiv \{\Psi^{(n)} = \{\Psi^{(n,m)}; m = 0,1,2,\cdots\} \in \mathcal{H}^{(n)} | \exists N_0 (\Psi^{(n)}) \geq 0$ such that $\Psi^{(n,m)} = 0$ for all $m > N_0(\Psi^{(n)}), N_0(\Psi^{(n)})$ being some finite number, depending on the vector $\Psi^{(n)}\}$. $\overset{0}{\mathcal{H}}{}^{(n)}$ is thus the set of all "vectors with only finitely many nonvanishing meson components";

(b) $\overset{0}{D}{}^{(n)}$ = subset of $\overset{0}{\mathcal{H}}{}^{(n)}$ consisting of all vectors $\Psi^{(n)} = \{\Psi^{(n,m)}\}$ with $\Psi^{(n,m)} \in C_0^\infty(\mathbb{R}^{3(n+m)})$ in the momentum variables, for all $m = 0,1,2,\cdots$.

Define $\mathcal{H}^{(n)}_{(\sigma)}$ as $\mathcal{H}^{(n)}$ for $\sigma < \infty$, and $\mathcal{H}^{(n)}_{(\sigma)} \equiv \mathcal{H}^{(n)}(R)$ for $\sigma = \infty$, where $\mathcal{H}^{(n)}(R)$ is the subspace of $\mathcal{H}^{(n)}$ consisting of functions which have support in each nucleon variable in a $\mathbb{R}^3$-ball of radius R. Similarly for $\mathcal{H}^{(n,m)}_{(\sigma)}$.

Set $E_R^{(n)} \equiv \{(q_1, q_2, \ldots, q_n | q_i \in \mathbb{R}^3; |q_i| \leq R, i = 1\cdots n\}$. Then $\mathcal{H}^{(n)}(R) \equiv \{\Phi^{(n)} \in \mathcal{H}^{(n)} | \text{supp } \Phi^{(n)} \subset E_R^{(n)}$ in the nucleon variables$\}$. Set also $\mathcal{H}^{(0)}(R) = \mathcal{H}^{(0)}$.

We have

Lemma 3.1a: For any $(0 \leq \sigma < \infty,\ 0 \leq \sigma' \leq \infty)$ there exists a number $\Lambda(\sigma,\sigma') > 0$ such that for all $|\lambda| \leq \Lambda$ (σ,σ') all operators introduced above [(3.4) to (3.9)] are defined on $D^{(n);\alpha}$ for some $\alpha > 0$ (dense in $\mathcal{H}^{(n)}$) and map $D^{(n);\alpha}$ into some $D^{(n);\beta}, 0 < \beta < \alpha$.

In particular:

(i) The mass renormalization $\hat{M}_\sigma(\sigma')$ is a bounded operator from $\mathcal{H}^{(n)}$ into $\mathcal{H}^{(n)}$. The estimates (2.5), (2.6), (2.7) of Lemma 2.1 hold with $\hat{M}_\sigma$ replaced by $\hat{M}_\sigma(\sigma')$ and V'_σ by $\lambda V_\sigma + \hat{M}_\sigma(\sigma')$.

(ii) The dressing operator $T_\sigma(\sigma')$ is defined on $D^{(n);\alpha}$ for every $\alpha > \frac{1}{2}$ ln2, maps $D^{(n);\alpha}$ for $\alpha > (1/2)$ ln2 into $D^{(n);\beta}$ for any $0 < \beta < \alpha - \frac{1}{2}$ ln 2. Note that $T_\sigma(\sigma') = 1$ on $\mathcal{H}^{(0)}$.

(iii) The field strength renormalization $A_\sigma(\sigma')$ is a contraction from $\mathcal{H}^{(n)}$ into $\mathcal{H}^{(n)}$, with strictly positive norm. $A_\sigma(\sigma')$ is invertible and commutes with H_0 on $D(H_0) \cap \mathcal{H}^{(n)}$. Note that $A_\sigma(\sigma') = 1$ on $\mathcal{H}^{(0)}$.

Proof: (i) follows from Ref. 20a. Note however that Eckmann treats from the very beginning only the case without cut-off ($\sigma = \infty$) and has consequently to restrict the nucleon states to the subspace $\mathcal{H}^{(n)}(R)$ of $\mathcal{H}^{(n)}$. In fact, the basic estimate in Lemma 9 of Ref. 20a holds, when $\sigma = \infty$, only for functions $\varphi \in \mathcal{H}^{(1,0)}$ whose support is restricted to some ball $|p| \leq R$. If, however, one has a finite cut-off $\sigma < \infty$, then automatically the operator $\Gamma^\alpha(W^{(i)})$ entering Lemma 9 (Ref. 20a) "picks up only momenta $|p| \leq \sigma$" in φ, and thus the estimate in Lemma 9 (Ref. 20a) holds for all $\varphi \in \mathcal{H}^{(1,0)}$, without restriction on the support [$D(R)$ is replaced by $D(\sigma)$]. The Lemmata 10 through 12 in Ref. 20a hold also for all $\varphi \in \mathcal{H}^{(1,0)}$ (with R replaced by σ in all the constants entering the estimates). Note in particular the estimate (from Lemma 11 in Ref. 20a and its corollary)

$$\|M_{\sigma 2k}(\sigma')|_{\mathcal{H}^{(n)}}\| \leq |\lambda|^{2k} n C(\sigma,\sigma')^{2^k}, \tag{3.11}$$

where $C(\sigma,\sigma')$ is some constant, depending on σ,σ' but not on k,λ,n. Points (ii) and (iii) are particular cases of the detailed statements of the following Lemma 3.1b.

To state this lemma we need the following notations: Let $e^{N_a\gamma}$ for any number γ be the operator in $\mathcal{H}$ whose restriction to every $\mathcal{H}^{(n,m)}$ is the operator multiplication by $e^{m\gamma}$. Let for any two Banach spaces B_1, B_2 be $\mathcal{L}(B_1,B_2)$ the set of all bounded operators from B_1 to B_2. Call $\|L\|_{B_1B_2}$ the norm of an operator $L \in \mathcal{L}(B_1,B_2)$. In the case where $B_1 = D^{(n);\alpha}, B_2 = D^{(n);\beta}$ we shall use the shorter notation $\|L\|_{\alpha\beta} \equiv \|L\|_{D^{(n);\alpha},D^{(n);\beta}}$.

Lemma 3.1b: For any $\{(\sigma,\sigma')|\ 0 \leq \sigma \leq \infty, 0 \leq \sigma' \leq \infty, \sigma' = R$ for $\sigma = \infty\}$ there exists a number $\lambda_1(\sigma,\sigma') > 0$ such that for all $|\lambda| \leq \lambda_1$ the following properties hold:

(i) If $Z_\sigma(\sigma')$ stands for $\Gamma Y_\sigma(\sigma'), Y_\sigma(\sigma'), T_\sigma(\sigma')$, where $Y_\sigma(\sigma')$ is any of the operators $S_{\sigma\nu,i}(\sigma'), \sum_\nu S_{\sigma\nu,i}(\sigma'), \sum_i S_{\sigma\nu,i}(\sigma'), Q_\sigma(\sigma'), Q_\sigma(\sigma')^*$, then $Z_\sigma(\sigma') \exp(-N_a \frac{1}{2} \ln 2) \in \mathcal{L}(\mathcal{H}^{(n)};\mathcal{H}^{(n)})$. Moreover, $Z_\sigma(\sigma')$ is defined on $D^{(n);\alpha}$ for all $\alpha > \frac{1}{2}$ ln 2 and $Z_\sigma(\sigma') \in \mathcal{L}_{\alpha\beta}$, for all $\alpha > \frac{1}{2}$ ln 2, $0 < \beta < \alpha - \frac{1}{2}$ ln 2. As a consequence we have also $Z_\sigma(\sigma') \in \mathcal{L}(D^{(n);\alpha};\mathcal{H}^{(n)})$.

(ii) Let, for $\sigma < \infty, X_\sigma$ be any of the operators $\lambda V^a_\sigma, \lambda V^c_\sigma, M_{\sigma 2k}, \sum_k M_{\sigma 2k}$. Then $X_\sigma \Gamma Y_\sigma, (\Gamma Y_\sigma)X_\sigma$ and $X_\sigma T_\sigma, T_\sigma X_\sigma$ are defined on $D^{(n);\alpha}$ for any $\alpha > \frac{1}{2}$ ln 2. [These are, by the convention made above, defined as equal to the corresponding quantities with labels $(\sigma,\sigma') = (\sigma,\infty)$].

(iii) For $\sigma < \infty, H_0\Gamma Y_\sigma$ and $\Gamma Y_\sigma H_0$ are defined on $D^{(n);\alpha} \cap D(H_0)$ for $\alpha > \frac{1}{2}$ ln 2 and $[H_0, \Gamma Y_\sigma] = Y_\sigma$ on these subsets. $H_0 T_\sigma, T_\sigma H_0$ and: $Q_\sigma T_\sigma$: are also defined on the same subsets $D^{(n);\alpha} \cap D(H_0)$, $\alpha > \frac{1}{2}$ ln 2 and $[H_0, T_\sigma] = :Q_\sigma T_\sigma:$ on these subsets.

(iv) For $\sigma < \infty$ we have on $D^{(n);\alpha} \cap D(H_0)$ with $\alpha > \frac{1}{2}$ ln 2:

$$\begin{aligned}\hat{H}_\sigma T_\sigma = T_\sigma H_0 + :T_\sigma\{&Q_\sigma + \lambda V^c_\sigma + \lambda V^a_\sigma + \hat{M}_\sigma + \lambda : V^a_\sigma \underbrace{\Gamma Q_\sigma}_{1} \\ &+ \lambda : V^a_\sigma \underbrace{\Gamma(Q_\sigma)}_{2} : + \frac{\lambda}{2} : V^a_\sigma \underbrace{\ \ }_{2} : (\Gamma(Q_\sigma))^2 : + \lambda : V^c_\sigma \underbrace{\Gamma(Q_\sigma)}_{1} : \\ &+ : \hat{M}_\sigma \underbrace{\Gamma(Q_\sigma)}_{1} :\} : = T_\sigma H_0 + :T_\sigma\{\lambda V^a_\sigma + \lambda V^a_\sigma \underbrace{\Gamma(Q_\sigma)}_{1} \\ &+ \frac{\lambda}{2} V^a_\sigma \underbrace{\ \ }_{2} : (\Gamma(Q_\sigma))^2 : . \end{aligned} \tag{3.12}$$

Note that $V^a_\sigma \underbrace{\ \ }_{2} : (\Gamma(Q_\sigma))^2 : = V^a_\sigma \angle (\Gamma(Q_\sigma))^2$ in Friedrichs notation (Ref. 22).

(v) The strong limit of $E_\sigma(R)$ for $\sigma \to \infty$, where $E_\sigma(R)$ stands for $\hat{M}'_\sigma(R)$ or $\Gamma(Q_\sigma(R))$ or $T_\sigma(R)$ or $A_\sigma(R)$ or $\nu_\sigma(R)$, exists on $D'^{(n)}$ and is equal $E_\infty(R)$. We shall denote $E_\infty(R)$ simply by E_∞, $D'^{(n)} \equiv \cup D^{(n);\alpha}$, $\alpha > \frac{1}{2}$ ln 2.

(vi) For $\sigma < \infty$ the operator $H^F_\sigma(R) \equiv H_0 + \lambda V_\sigma + M_{\sigma 2}(R)$ is self-adjoint and bounded from below in each sector $\mathcal{H}^{(n)}$, with domain $D(H^F_\sigma(R)) \cap \mathcal{H}^{(n)} = D(H_0) \cap \mathcal{H}^{(n)}$.

Set $\hat{H}'_\sigma(R) \equiv H^F_\sigma(R) + \hat{M}'_\sigma(R)$. The Born series expansion of $(z - \hat{H}'_\sigma(R))^{-1}$ converges in norm, for $\sigma \to \infty$, in $\mathcal{H}^{(n)}$ and for all Rez suitably negative, to $(z - \hat{H}_\infty)^{-1}$, where $\hat{H}_\infty$ is a self-adjoint lower bounded operator in each sector $\mathcal{H}^{(n)}$.

(vii) $\hat{H}_\infty$ is essentially self-adjoint on the subset $\overline{T^s_\infty \overset{0}{D}{}^{(n)}}$ of $\overline{T^s_\infty \overset{0}{D}{}^{(n)}} \equiv \hat{\mathcal{H}}^{s(n)}(R)$, where $\overset{0}{D}{}^{(n)}$ is as in Remark 3.1. $T^s_\infty \equiv$ s-$\lim_{\sigma\to\infty}$: exp$[-\lambda\Gamma(V^c_\sigma(R))]$: and $-$ means closure. On $\hat{\mathcal{H}}^{s(n)}(R)$ one has $\hat{H}_\infty = H^F_\infty + \hat{M}'_\infty$, where H^F_∞ is, in each $\mathcal{H}^{(n)}$, the self-adjoint, lower bounded operator such that n-$\lim_{\sigma\to\infty}$ $(z - H^F_\sigma(R))^{-1} = (z - H^F_\infty)^{-1}$.

(viii) For every $\phi \in D'^{(n)} \cap \mathcal{H}^{(n)}(R)$, $T_\infty\phi$ is in the domain of $\hat{H}_\infty$

and

$$\begin{aligned}\hat{H}_\infty T_\infty\phi = (H^F_\infty + \hat{M}'_\infty)T_\infty\phi &= T_\infty H_0\phi \\ &+ :T_\infty\{\lambda V^a + \lambda V^a \underbrace{\Gamma(Q_\infty)}_{1} + \tfrac{1}{2}\lambda V^a \angle : (\Gamma(Q_\infty))^2 :\} : \phi,\end{aligned} \tag{3.13}$$

where

$$V^a \angle : (\Gamma(Q_\infty))^2 : = V^a \underbrace{\ \ }_{2} : (\Gamma(Q_\infty))^2 : .$$

(ix) A_∞ has the same properties as A_σ in Lemma 3.1a, (iii).

Proof of Lemma 3.1b:

(i) That $T_\sigma(\sigma')$ is defined on $D^{(n);\alpha}$, $\alpha > (\frac{1}{2})\ln 2$ follows from the estimate, valid for all $\sigma < \infty$, $|\lambda| \leq \Lambda(\sigma)$ for some constant $\Lambda(\sigma) > 0$:

$$\|(T_\sigma(\sigma')\Psi^{(n,m)})^{(n,i+m)}\|^2 \leq 2^m C_\sigma'^i (i!)^{-1+2\epsilon} \|\Psi^{(n,m)}\|^2, \quad (3.14)$$

for any $\Psi^{(n,m)} \in \mathcal{K}^{(n,m)}$, where ϵ is any number satisfying $0 < \epsilon < \frac{1}{4}$, C'_σ a constant which depends (possibly) on σ, λ, n, ϵ but is independent of m, i. This estimate has been derived by Eckmann (see the end of the proof of Theorem 8 in Ref. 20a). Similar estimates show that $Z_\sigma(\sigma')$ is defined $D^{(n);\alpha}$. Moreover, $T_\sigma(\sigma') \in \mathcal{L}_{\alpha\beta}$ follows easily from (3.14), and similarly $Z_\sigma(\sigma') \in \mathcal{L}_{\alpha\beta}$.

(ii) Since we know from (i) that $\Gamma Y_\sigma, T_\sigma$ map $D^{(n);\alpha}$ into $D^{(n);\beta}$, $\beta < \alpha - (\frac{1}{2})\ln 2, \alpha > (\frac{1}{2})\ln 2$, this point is trivial, since $M_{\sigma 2k}, \hat{M}_\sigma$ are bounded on $\mathcal{K}^{(n)}$ and map $\mathcal{K}^{(n,m)}$ into itself and V^a_σ, V^c_σ map a dense subset of $\mathcal{K}^{(n,m)}$ into $\mathcal{K}^{(n,m-1)}$ resp. $\mathcal{K}^{(n,m+1)}$, hence any $D^{(n);\beta}$ into $D^{(n);\beta'}$, $\beta' < \beta$.

(iii) Clearly $\Gamma Y_\sigma H_0$ is defined on $D^{(n);\alpha} \cap D(H_0)$ for any $\alpha > (\frac{1}{2})\ln 2$. The operator $[H_0, \Gamma Y_\sigma]$ acts on each $\mathcal{K}^{(n,m)}$ exactly as Y_σ. Since from (i) we know that Y_σ is defined on $D^{(n);\alpha}$, we have that $\Gamma Y_\sigma H_0$ and $[H_0, \Gamma Y_\sigma]$ are both defined on $D^{(n);\alpha} \cap D(H_0)$. Hence also $H_0 \Gamma Y_\sigma$ and $T_\sigma H_0$ are defined on the same domain, the latter because of (i) and the fact that H_0 conserves the number of particles.

Moreover, $[H_0, :(\Gamma(Q_\sigma))^j:]$ is in $\mathcal{L}_{\alpha\beta}$, as seen from the estimates in (i) [the operation ad H_0 modifies just the kernel of $:(\Gamma(Q_\sigma))^j:$ and this does not change the form of the estimate, since $\sigma < \infty$.

Since $:(\Gamma(Q_\sigma))^j:\varphi = 0$ for all $\varphi \in D^{(n);\alpha}, j > n$, this implies also that $[H_0, T_\sigma]$ is defined on $D^{(n);\alpha}$. This, together with above observation $D(T_\sigma H_0) \supset D^{(n);\alpha} \cap D(H_0)$, implies that $H_0 T_\sigma$ is also defined on $D^{(n);\alpha} \cap D(H_0)$. Moreover, one sees that $:Q_\sigma T_\sigma:$ is defined on $D^{(n);\alpha}$, $\alpha > \frac{1}{2}\ln 2$, since, on $\varphi \in D^{(n);\alpha}$, it gives a sum of $n+1$ terms $(j!)^{-1}:Q_\sigma(:(\Gamma(Q_\sigma))^j:):\varphi, (j \leq n)$, which are all in $D^{(n);\alpha}$ [as seen by computing, for $\Psi \in \mathcal{K}^{(n,m)}$,

$$(:Q_\sigma(:(\Gamma(Q_\sigma))^j:):\Psi)^{(n,m+i)} = \left(\sum_{k+\sum_t l_t = i} \sum_\nu S_{\sigma\nu,k} \prod_{k=1}^{j} \sum_\nu S_{\sigma\nu,l_t} \Psi\right)^{(n,m+i)}].$$

By computing we have now $H_0 T_\sigma = T_\sigma H_0 + :T_\sigma Q_\sigma:$ on $D^{(n);\alpha} \cap D(H_0)$.

(iv) We have already shown that the first two terms on the right hand side of (3.12) are defined on $D^{(n);\alpha} \cap D(H_0)$. The others are also defined on the same set, as seen similarly, by first computing an estimate on the kernel-norm of each term in the $\{\ \}$ bracket, along the lines followed for estimating the kernels of $\Gamma(Q_\sigma)$. The equality (3.12) follows then by the standard Wick ordered expression of the form $A: \exp B:, A, B$, being Wick monomials (see, e.g., Lemma 1.2, Lemma 1.3 in K. Hepp, Ref. 21).

(v) The existence of the limit and the fact that the quantities E_∞ are defined on $D'^{(n)}$ is a consequence of the same estimates of Eckmann which implied (i) above (in particular, Ref. 20a, p. 263).

(vi) One first shows, proceding as in the proof of Theorem 20 in Ref. 20b, that the Born series for $(z - H^F_\sigma(R))^{-1}$ converges in norm for suitable negative Re z to a limit which is the resolvent of an operator H^F_∞, which is self-adjoint and bounded below on each $\mathcal{K}^{(n)}$. [The only difference with Theorem 20 in Ref. 20b is that M_σ has to be replaced by $M_{\sigma 2}$ in the definition of $(V^a_\sigma R_0(z) V^c_\sigma)_{\text{ren}}$.] Since the sum $\hat{M}'_\sigma(R)$ of all renormalization terms of order > 2 has a norm which is uniformly bounded[20a] in σ, the estimates (18), (19) in Ref. 20b are still valid if $(V^a_\sigma R_0(z) V^c_\sigma)_{\text{ren}}$ is defined with $M_{\sigma 2} + \hat{M}'_\sigma(R)$ instead of $M_{\sigma 2}$. Hence the Born series for $(z - \hat{H}'_\sigma(R))^{-1}$ also converges in norm. One shows as in Ref. 20b, that

$$\text{n-}\lim_{\sigma\to\infty} (z - \hat{H}'_\sigma(R))^{-1} = (z - \hat{H}_\infty)^{-1},$$

where $\hat{H}_\infty$ is self-adjoint and lower bounded in each $\mathcal{K}^{(n)}$.

(vii) It suffices to show that H^F_∞ is essentially self-adjoint on $T^s_\infty \overset{0}{D}{}^{(n)}$ proceding as in Theorem 19, 24 of Ref. 20b. As in the same proof [or also as in (v) above], one proves also $\text{s-}\lim_{\sigma\to\infty} H^F_\sigma(R) T^s_\sigma(R)\phi = H^F_\infty T^s_\infty \phi$. Since $\hat{M}'_\sigma(\sigma')$ is uniformly bounded in σ, we have on the other hand

$$\text{s-}\lim_{\sigma\to\infty} \hat{M}'_\sigma(R) T^s_\sigma(R)\phi = \hat{M}'_\infty T^s_\infty \phi,$$

and therefore

$$\text{s-}\lim_{\sigma\to\infty} \hat{H}'_\sigma(R) T^s_\sigma(R)\phi = H^F_\infty T^s_\infty \phi + \hat{M}'_\infty T^s_\infty \phi.$$

But

$$\begin{aligned} T^s_\infty \phi &= \text{s-}\lim_{\sigma\to\infty} T^s_\sigma(R)\phi \\ &= \text{s-}\lim_{\sigma\to\infty} (z - \hat{H}'_\sigma(R))^{-1}(z - \hat{H}'_\sigma(R)) T^s_\sigma(R)\phi \\ &= (z - \hat{H}_\infty)^{-1}(z T^s_\infty \phi - H^F_\infty T^s_\infty \phi - \hat{M}'_\infty T^s_\infty \phi). \end{aligned}$$

Hence $T^s_\infty \phi \in D(\hat{H}_\infty)$ and

$$\begin{aligned} \hat{H}_\infty T^s_\infty \phi &= \hat{H}_\infty (z - \hat{H}_\infty)^{-1}(z T^s_\infty \phi - H^F_\infty T^s_\infty \phi - \hat{M}'_\infty T^s_\infty \phi) \\ &= [z(z - \hat{H}_\infty)^{-1} - 1](z T^s_\infty \phi - H^F_\infty T^s_\infty \phi - \hat{M}'_\infty T^s_\infty \phi) \\ &= \text{s-}\lim_{\sigma\to\infty} [z(z - \hat{H}_\infty)^{-1} - 1](z T^s_\sigma(R)\phi - \hat{H}'_\sigma(R) T^s_\sigma(R)\phi) \\ &= \text{s-}\lim_{\sigma\to\infty} \hat{H}'_\sigma(R) T^s_\sigma(R)\phi = H^F_\infty T^s_\infty \phi + \hat{M}'_\infty T^s_\infty \phi. \end{aligned}$$

From this it follows $\hat{H}_\infty = H^F_\infty + \hat{M}'_\infty$ on $T^s_\infty \overset{0}{D}{}^{(n)}$, which is what we wanted.

(viii) One has from (3.12), for any $\phi \in \mathcal{K}^{(n,m)} \cap \mathcal{K}^{(n)}(R) \cap \overset{0}{D}{}^{(n)}$, using the facts that $T_\sigma(R) = T_\sigma$ on $\mathcal{K}^{(n)}(R)$ and that the right hand side of (3.12) on $\mathcal{K}^{(n)}(R)$ is equal to the same expression with an R cut-off in all operators except H_0:

$$\begin{aligned} \hat{H}_\sigma T_\sigma(R)\phi = T_\sigma(R) H_0 \phi \\ + :T_\sigma(R)\{\lambda V^a_\sigma(R) + \lambda V^a_\sigma(R) \underbrace{\Gamma(Q_\sigma(R))}_{1} \\ + \frac{\lambda}{2} V^a_\sigma(R) \underbrace{:(\Gamma(Q_\sigma(R)))^2:}_{2}\}\phi. \end{aligned}$$

We now study the limit $\sigma \to \infty$. We claim that the right-hand side converges strongly to the right-hand side of the similar expression in point (viii) of Lemma 3.1b. This is proved using again the estimates of Ref. 20.

Let us study as an example the term

$$\lambda :T_\sigma(R) V^a_\sigma(R):\phi = \lambda T_\sigma(R) V^a_\sigma(R)\phi - \lambda T_\sigma(R) \underbrace{V^a_\sigma(R)}_{1}\phi.$$

All expressions are well defined because of Lemma 3.1b(i)-(v), above. To see that $T_\sigma(R) V^a_\sigma(R)\phi \to T_\infty V^a(R)\phi$ strongly as $\sigma \to \infty$, it suffices to use the fact that $\|(T_\sigma(R)$

$-T_\infty(R))e^{-N_a(1/2)\ln 2}\| \to 0$ and $\|(V_\sigma^a(R) - V^a(R))\phi\| \to 0$ as $\sigma \to \infty$ (see Lemma 3.1b above).

Similarly one shows $T_\sigma(R)\underset{1}{\smile}V_\sigma^a(R)\phi \to T_\infty(R)\underset{1}{\smile}V^a(R)\phi$, which then implies: $T_\sigma(R)V_\sigma^a(R):\phi \to :T_\infty(R)V^a(R):\phi$. In the same way one proves the convergence of the other terms to the corresponding limits. Since $\hat{M}'_\sigma T_\sigma(R)\phi \to \hat{M}'_\infty T_\infty\phi$, we have that $\underset{\sigma\to\infty}{\text{s-lim}}\, H_\sigma^F T_\sigma(R)\phi$ also exists. This together with the norm convergence $(z - H_\sigma^F)^{-1} \to (z - H_\infty^F)^{-1}$ for Re z sufficiently negative (as in Ref. 20b) gives, as in the proof of (iii) above, $\underset{\sigma\to\infty}{\text{s-lim}}\, H_\sigma^F T_\sigma(R)\phi = H_\infty^F T_\infty\phi$. This concludes the proof of (viii).

Point (ix) is proved as in Lemma 3.1a. ∎

Theorem 3.1: For every $0 \le \sigma < \infty$ there exists a number $\lambda_0(\sigma) > 0$ and for $\sigma = \infty$ and any given number $0 \le R < \infty$ there exists a number $\lambda_0'(R) > 0$ such that for all $|\lambda| \le \lambda_0(\sigma)$ resp. $|\lambda| \le \lambda_0'(R)$ we have the following properties:

(i) The renormalized Hamiltonian $\hat{H}_\sigma$, defined for $\sigma < \infty$ by $H_0 + \lambda V_\sigma + \hat{M}_\sigma$ and for $\sigma = \infty$ by Lemma 3.1b, (vi), is self-adjoint and lower bounded in $\mathcal{H}^{(n)}$. $\hat{H}_\sigma$ extends in the natural way to a self-adjoint operator in $\mathcal{H}$, which is bounded from below on every subspace of $\mathcal{H}$ with a finite number of nucleons.[42] For $\sigma < \infty$, $\hat{H}_\sigma$ is defined on the domain $D(H_0) \cap \mathcal{H}^{(n)}$, dense in $\mathcal{H}^{(n)}$. For $\sigma = \infty$, $\hat{H}_\sigma$ has a domain containing in particular $T_\infty(\overset{0}{D}{}^{(n)} \cap \mathcal{H}^{(n)}(R))$.

(ii) $\hat{H}_\sigma$ has the correct relativistic one particle spectrum. One has $\hat{H}_\sigma T_\sigma A_\sigma = T_\sigma A_\sigma H_0$ on $(\mathcal{H}^{(1,0)} \cup \mathcal{H}^{(0)}) \cap D(H_0)$, i.e., for any g, h such that $\omega g \in L_2(\mathbb{R}^3), \mu h \in L_2(\mathbb{R}^3)$:

$$\hat{H}_\sigma T_\sigma A_\sigma b^*(g)\Omega_0 = T_\sigma A_\sigma H_0 b^*(g)\Omega_0 = T_\sigma A_\sigma b^*(\omega g)\Omega_0,$$

$$\hat{H}_\sigma a^*(h)\Omega_0 = H_0 a^*(h)\Omega_0 = a^*(\mu h)\Omega_0.$$

The "physical one particle states" $T_\sigma A_\sigma b^*(g)\Omega_0, a^*(h)\Omega_0$ are correctly normalized:

$$\|T_\sigma A_\sigma b^*(g)\Omega_0\| = 1, \quad \|a^*(h)\Omega_0\| = 1,$$

for

$$\|g\| = 1, \quad \|h\| = 1.$$

On $\mathcal{H}^{(0)} \cap D(H_0)$, $\hat{H}_\sigma$ coincides with H_0.

Corollary: For $|\lambda| \le \lambda_1(\sigma, R), \lambda_1$ as in Lemma 3.1b, we have $\underset{\sigma\to\infty}{\text{s-lim}}\, e^{it\hat{H}'_\sigma(R)} = e^{it\hat{H}}\infty$ on $\mathcal{H}^{(n)}$.

Remark 3.1: $T_\sigma A_\sigma$ maps isometrically $\mathcal{H}^{(1,0)}$ (all states of a single bare nucleon) onto a closed subspace $\mathcal{H}^{(1,0)}_{(\sigma)\text{phys}}$ of $\mathcal{H}^{(1)}_{(\sigma)\text{phys}} \subset \mathcal{H}^{(1)}$ (all states consisting of a single physical nucleon). The restriction of $T_\sigma A_\sigma$ to $\mathcal{H}^{(1)}$ is precisely a partial isometry with initial domain $\mathcal{H}^{(1,0)}$ and final domain $\mathcal{H}^{(1,0)}_{(\sigma)\text{phys}}$. $T_\sigma A_\sigma$ coincides on $\mathcal{H}^{(1,0)}$ with the wave operators $\tilde{\Omega}_\sigma^\pm$ defined below (Theorem 3.4).

Proof of Theorem 3.1: Theorem 3.1 is an immediate consequence of the Lemmata 3.1a and b. The Corollary follows from Lemma 3.1, (vi) and Trotter's convergence theorem. ∎

We proceed in the following to the construction of creation–annihilation operators for physical nucleons. As in Sec. 2 we shall usually restrict all operators to the n nucleon sector $\mathcal{H}^{(n)}$ dropping whenever not necessary for the understanding the label $|_n$ we used some times before for indicating these restrictions. Physical operators will be distinguished from bare ones by a "hat" $\hat{}$. We make the convention of denoting always by an upper prime $'$ and a lower σ label the function obtained from a given element $g \in L_2(\mathbb{R}^3)$ by multiplying it with the normalization function $\nu_\sigma^{-1/2}(\cdot)$, ν_σ being as in (3.9), (3.10): thus we will write

$$g'_\sigma(q) \equiv g(q)\cdot\nu_\sigma^{-1/2}(q). \tag{3.15}$$

Define

$$D_{(\sigma)}^{(n);\alpha} \equiv D^{(n);\alpha} \cap \mathcal{H}_{(\sigma)}^{(n)}, \qquad D_{(\sigma)}^{\prime(n)} = \bigcup_{\alpha > (1/2)\ln 2} D_{(\sigma)}^{(n);\alpha}$$

Theorem 3.2a: Let, for $0 \le \sigma < \infty, \lambda_0(\sigma) > 0$ and, for $\sigma = \infty, \lambda_0'(R) > 0$ be as in Theorem 3.1. Set $\hat{\Lambda} = \lambda_0(\sigma)$ for $\sigma < \infty$ and $\hat{\Lambda} = \lambda_0'(R)$ for $\sigma = \infty$. Then for all $|\lambda| \le \hat{\Lambda}$ and all $0 \le \sigma \le \infty$ we have: $\hat{H}_\sigma$ and $\Gamma(Q_\sigma)$ are defined by Theorem 3.1 and for all $h \in L_2(\mathbb{R}^3), \|h\| = 1$ the following properties hold:

(i) Set

$$\overset{0}{B}{}^*(h'_\sigma) \equiv b^*(h'_\sigma) + \Gamma(Q_\sigma)\underset{1}{\smile}b^*(h'_\sigma), \tag{3.16}$$

where b^* is the creator of bare nucleons defined in Sec. 1. $\overset{0}{B}{}^*(h'_\sigma)$ is defined on the dense subset $D_{(\sigma)}^{\prime(n)}$ of $\mathcal{H}_{(\sigma)}^{(n)}$ and belongs to $\mathcal{L}(D_{(\sigma)}^{(n);\alpha}; D^{(n+1);\alpha'})$, for any $\alpha > \frac{1}{2}\ln 2$, $0 < \alpha' < \alpha - \frac{1}{2}\ln 2$.

Let $\overset{0}{B}(\bar{h}'_\sigma)$ be the adjoint of $\overset{0}{B}{}^*(h'_\sigma)$. $\overset{0}{B}{}^*(\bar{h}'_\sigma)$ is defined on the dense subset $D_{(\sigma)}^{\prime(n+1)}$ and belongs to $\mathcal{L}(D_{(\sigma)}^{(n+1);\alpha}; D^{(n);\alpha'})$, for any $\alpha > \frac{1}{2}\ln 2, 0 < \alpha' < \alpha - \frac{1}{2}\ln 2$. On the dense subset $D_{(\sigma)}^{(n+1);\alpha}, \alpha > \frac{1}{2}\ln 2$ of $\mathcal{H}_{(\sigma)}^{(n+1)}$ we have:

$$\overset{0}{B}(\bar{h}'_\sigma) = (\overset{0}{B}(h'_\sigma))^* = b(\bar{h}'_\sigma) + b(\bar{h}'_\sigma)\underset{1}{\smile}(\Gamma(Q_\sigma))^*. \tag{3.17}$$

$\overset{0}{B}{}^*(h'_\sigma)$ and $\overset{0}{B}(\bar{h}'_\sigma)$ are closable and we shall call their closures $B^*(h'_\sigma)$ resp. $B(\bar{h}'_\sigma)$.

These are the creation resp. annihilation operators for dressed nucleons.

(ii) We have

$$T_\sigma A_\sigma \prod_{i=1}^{n} b^*(h_i)\phi = \prod_{i=1}^{n} B^*(h'_{\sigma i})\phi, \tag{3.18}$$

for any $h_i \in L_2(\mathbb{R}^3), \|h_i\| = 1$, supp $h_i \in E_R^{(1)}, \phi \in D^{n(0)} \equiv \cup_{\alpha > (n/2)\ln 2} D^{(0);\alpha}$.

Thus $T_\sigma A_\sigma$ maps states in $\mathcal{H}_{(\sigma)}^{(n)}$ with n bare nucleons and arbitrarily many (even infinitely many) mesons into states with n physical nucleons and arbitrarily many mesons.

Proof of Theorem 3.2a: The case $\sigma < \infty$ is proved as follows:

(i) One has $b^*(h'_\sigma) \in \mathcal{L}(\mathcal{H}^{(n)}; \mathcal{H}^{(n+1)})$ and in particular $b^*(h'_\sigma) \in \mathcal{L}(D^{(n);\beta}; D^{(n+1);\beta})$ for all $\beta > 0$. $\Gamma(Q_\sigma)b^*(h'_\sigma)$ and $b^*(h'_\sigma)$ are both in $\mathcal{L}(D^{(n);\alpha}; D^{(n+1);\alpha'}), \alpha > \frac{1}{2}\ln 2, 0 < \alpha' < \alpha - \frac{1}{2}\ln 2$, and $\Gamma(Q_\sigma) \in \mathcal{L}(D^{(n+1);\alpha}; D^{(n+1);\alpha'})$, as proved in Lemma 3.1b, (i).

Hence $\Gamma(Q_\sigma)\underset{1}{\smile}b^*(h'_\sigma) = [\Gamma(Q_\sigma), b^*(h'_\sigma)]$ and $\overset{0}{B}{}^*$ are also in $\mathcal{L}(D^{(n);\alpha}; D^{(n+1);\alpha'})$. The adjoint of B^* is defined on $D^{(n+1);\alpha}$, since those of b^* and $\Gamma(Q_\sigma)\underset{1}{\smile}b^*(h'_\sigma)$ exist on this set and are equal to b and $[b(h'_\sigma), (\Gamma(Q_\sigma))^*]$, respectively (from Lemma 3.1b).

The case $\sigma = \infty$ is proven in the same way, using the corresponding statements in Lemma 3.1b. ■

Theorem 3.2b: Let $0 \le \sigma < \infty$, $|\lambda| < \lambda_0(\sigma)$. Set $\hat{b}^{\#}_{\sigma,t}(h'_\sigma) \equiv e^{it\hat{H}_\sigma} B^{\#}(e^{\mp it\omega} h'_\sigma) e^{-it\hat{H}_\sigma}$. Then:

(i) The $\hat{b}^{\#}_{\sigma,t}(h'_\sigma)$ are closed operators belonging to $\mathcal{L}(D^{(n);\alpha}; D^{(n\pm1);\alpha'})$ for any $\alpha > \frac{1}{2}\ln 2, 0 < \alpha' < \alpha - \frac{1}{2}\ln 2$.

They are the "adjusted dressed nucleon creation and annihilation operators" (in short "dressed nucleon fields") in the Heisenberg picture.

(ii) The following commutation relations hold for any $f, g \in L_2(\mathbb{R}^3)$:

$$[\hat{b}^{*}_{\sigma,t}(f'_\sigma), \hat{b}^{*}_{\sigma,t}(g'_\sigma)] = 0 = [b_{\sigma,t}(f'_\sigma), b_{\sigma,t}(g'_\sigma)] \tag{3.19a}$$

on

$$D''^{(n)} \equiv \bigcup_{\alpha > \ln 2} D^{(n);\alpha}$$

and for expectation values with respect to the vacuum Ω_0

$$(\Omega_0, [\hat{b}_{\sigma,t}(\bar{f}'_\sigma), \hat{b}^{*}_{\sigma,t}(g'_\sigma)]\Omega_0) = (f, g), \tag{3.19b}$$

where the primed quantities f'_σ, g'_σ are obtained from f, g in the same way as h'_σ was obtained from h [see (i) above] and $(,)$ is the scalar product in L_2.

(iii) Let $\hat{\Lambda}$ be as in Theorem 3.2a. Set, for $|\lambda| \le \hat{\Lambda}$ and all $\sigma \le \infty$,

$$\hat{b}^{\#}_{\sigma,t}(R; h'_\sigma(R)) \equiv e^{it\hat{H}'_\sigma(R)} B^{\#}(e^{\mp it\omega} h'_\sigma(R)) e^{-it\hat{H}'_\sigma(R)},$$

$$\hat{b}^{\#}_{\infty,t}(h'_\infty) \equiv e^{it\hat{H}_\infty} B^{\#}(e^{\mp it\omega} h'_\infty) e^{-it\hat{H}_\infty}.$$

Then we have that $D'^{(0)} \cup \mathcal{K}^{(1,0)}$ is contained in the domain of $\hat{b}^{\#}_{\sigma,t}(h'_\sigma)$ and

$$\hat{b}^{*}_{\sigma,t}(h'_\sigma) T_\sigma A_\sigma \phi = \Omega_\sigma(t) b^*(h)\phi,$$

where

$$\Omega_\sigma(t) \equiv e^{it\hat{H}_\sigma} T_\sigma A_\sigma e^{-itH_0}.$$

Moreover,

$$\hat{b}^{*}_{\infty,t}(h'_\infty) T_\infty A_\infty \phi = \operatorname{s-lim}_{\sigma\to\infty} \hat{b}^{*}_{\sigma,t}(R; h'_\sigma(R)) T_\sigma(R) A_\sigma(R)\phi$$
$$= \operatorname{s-lim}_{\sigma\to\infty} \Omega_\sigma(R; t) b^*(h)\phi, \tag{3.20}$$

where

$$\Omega_\sigma(R; t) \equiv e^{it\hat{H}'_\sigma(R)} T_\sigma(R) A_\sigma(R) e^{-itH_0}.$$

Note that $T_\sigma A_\sigma \phi = T_\sigma(R) A_\sigma(R)\phi = \phi$ for $\phi \in D'^{(0)}$.

In particular, for $\phi = \Omega_0$ one has

$$\hat{b}^{*}_{\sigma,t}(h'_\sigma)\Omega_0 = T_\sigma A_\sigma b^*(h)\Omega_0 = \Omega_\sigma(t) b^*(h)\Omega_0.$$

Furthermore,

$$\hat{b}_{\sigma,t}(h'_\sigma)\Phi = \hat{b}_{\sigma,t}(R; h'_\sigma(R))\Phi = 0,$$

for all $\Phi \in \mathcal{K}^{(0)}$, hence in particular for $\Phi = \Omega_0$.

(iv) For λ and σ as in (iii) set, for any $g \in L_2(\mathbb{R}^3)$, as in Sec. 2,

$$\hat{a}^{\#}_{\sigma,t}(g) \equiv e^{it\hat{H}_\sigma} a^{\#}(e^{\mp it\mu} g) e^{-it\hat{H}_\sigma}$$

and

$$\hat{a}^{\#}_{\sigma,t}(R; g) \equiv e^{it\hat{H}'_\sigma(R)} a^{\#}(e^{\mp it\mu} g) e^{-it\hat{H}'_\sigma(R)}.$$

Then Theorem 2.2 holds for $\hat{a}^{\#}_{\sigma,t}(g)$, for all $\sigma \le \infty$.

Moreover, for any $\Psi \in D(H_0^{1/2})$: $\hat{a}^{\#}_{\infty,t}(h)\Psi = \operatorname{s-lim}_{\sigma\to\infty} \hat{a}^{\#}_{\sigma,t}(R; h)\Psi$. One has also $\hat{a}^{*}_{\sigma,t}(g) T_\sigma A_\sigma \phi = \Omega_\sigma(t) a^*(g)\phi$, for any $\phi \in \mathcal{K}^{(1,0)}$.

Note that in particular, for $\sigma < \infty$, $\hat{a}^{\#}_{\sigma,t}(g) \in \mathcal{L}(D^{(n);\alpha}; D^{(n);\alpha'})$, for any $0 < \alpha' < \alpha$. On $D'^{(n)}$ one has, for $\sigma < \infty$,

$$[\hat{b}^{*}_{\sigma,t}(f'_\sigma), \hat{a}^{*}_{\sigma,t}(g)] = 0, \tag{3.21}$$

for any $f, g \in L_2(\mathbb{R}^3)$.

(v) For $\sigma < \infty$, the product of k nucleon operators and m meson operators $(k, m = 0, 1, 2, \cdots)$

$$\prod_{j=1}^{k} \hat{b}^{\#}_{\sigma,t}(f'^{(j)}_\sigma) \prod_{i=1}^{m} \hat{a}^{\#}_{\sigma,t}(g^{(i)})$$

(with $\prod^0 \equiv 1$) is defined on $D^{k(n)} \equiv \cup_{\alpha > (k/2)\ln 2} D^{(n);\alpha}$ (and hence in particular on states with a finite number of particles). For $\sigma = \infty$ this is defined for all $f^{(j)} \in \mathcal{K}^{(1,0)}(R)$ on $D'^{(0)}$ and on $T_\infty A_\infty \mathcal{K}^{(1,0)}$. Moreover, one has, e.g., for all $\sigma \le \infty$,

$$\prod_{j=1}^{n} \hat{b}^{*}_{\sigma,t}(f'^{(j)}_\sigma) \prod_{i=1}^{m} \hat{a}^{*}_{\sigma,t}(g^{(i)})\Omega_0 = \Omega_\sigma(t) \prod_{j=1}^{n} b^*(f^{(j)}) \prod_{i=1}^{m} a^*(g^{(i)})\Omega_0. \tag{3.22}$$

[This is a state consisting of n physical nucleons and m mesons which we expect to converge for $t \to \pm\infty$ to "scattering states" $\Omega^{\pm}_\sigma \prod_j b^*(f^{(j)}) \prod_i a^*(g^{(i)})\Omega_0$, i.e., to states in the ranges of the wave operators.]

Remark 3.2: One sees immediately from the definition that

$$B^*(h'_\sigma) = \hat{b}^{*}_{\sigma,t=0}(h'_\sigma) = b^*(h'_\sigma) + \Gamma(Q_\sigma) \underset{1}{\smile} b^*(h'_\sigma)$$

is a sum of infinitely many terms with $0, 1, 2, \ldots, m, \cdots$ meson creators, i.e., is of the "operator form" $B^* = b^* + \sum_{i=1}^{\infty} b^*(a^*)^i$. This corresponds to the familiar picture of the dressed nucleon as an undressed one accompanied by a "cloud" of infinitely many mesons.

Remark 3.3: As can be expected the commutators $[\hat{b}_{\sigma,t}(f'_\sigma), \hat{b}^{*}_{\sigma,t}(g'_\sigma)]$ on general states are not c-numbers. We shall see that they become c-numbers in the limit $t \to \pm\infty$. As an example, one can easily compute, for $\sigma < \infty$, that on $\mathcal{K}^{(n,0)}$

$$[\hat{b}_{\sigma,0}(\bar{f}'_\sigma), \hat{b}^{*}_{\sigma,0}(g'_\sigma)] = [B(\bar{f}'_\sigma), B^*(g'_\sigma)]$$
$$= (f'_\sigma, g'_\sigma) + [b(\bar{f}'_\sigma), \Gamma(Q_\sigma) \underset{1}{\smile} b^*(g'_\sigma)]$$
$$+ b(\bar{f}'_\sigma) \underset{1}{\smile} (\Gamma(Q_\sigma))^* \Gamma(Q_\sigma) \underset{1}{\smile} b^*(g'_\sigma) \tag{3.23}$$

Proof of Theorem 3.2b:

(i) We first prove that $B^{\#}(e^{\mp it\omega} h'_\sigma) e^{-it\hat{H}_\sigma}$ maps $D^{(n);\alpha}$, $\alpha > \frac{1}{2}\ln 2$, into $D^{(n);\alpha'}$, $0 < \alpha' < \alpha - \frac{1}{2}\ln 2$. We shall write the proof only for the case of B^*, the other case being treated similarly. We need only to investigate the term $\Gamma(Q_\sigma) \underset{1}{\smile} b^*(e^{-it\omega} h'_\sigma) e^{-it\hat{H}_\sigma}$. Because of Theorem 3.2a, (i) this will be in $\mathcal{L}(D^{(n);\alpha}; D^{(n+1);\alpha'})$, $0 < \alpha' < \alpha - (1/2)\ln 2$ as soon as one can show that $e^{-it\hat{H}_\sigma}$ maps $D^{(n);\alpha}$, $\alpha > (1/2)\ln 2$, into $D^{(n);\beta}$ for all $0 < \beta < \alpha$. But this is true because of the following:

Lemma 3.2: Let $0 \le \sigma < \infty$ be given and let $\lambda_0(\sigma)$ and $V'_\sigma \equiv \lambda V_\sigma + \hat{M}_\sigma$ be as in Theorem 3.1. Then for all $|\lambda| \le \lambda_0(\sigma)$ the time-dependent Dyson series

$$e^{itH_0}\left[1 + i\int_0^t dt_1 e^{-it_1H_0}V'_\sigma e^{it_1H_0} + i^2\int_0^t dt_1\int_0^{t_1} dt_2 e^{-it_1H_0}V'_\sigma e^{i(t_1-t_2)H_0}V'_\sigma e^{it_2H_0} + \cdots\right] \tag{3.24}$$

converges in the uniform operator topology for bounded operators from $D^{(n);\alpha}$ to $D^{(n);\beta}$ for any $0 < \beta < \alpha$ and the sum is equal $e^{it\hat{H}_\sigma}$ on $D^{(n);\alpha}$.

Hence in particular $e^{it\hat{H}_\sigma} \in \mathcal{L}(D^{(n);\alpha}; D^{(n);\beta})$ and $e^{it\hat{H}_\sigma}$ is strongly differentiable on $D^{(n);\alpha} \cap D(H_0)$ in the $D^{(n);\beta}$-topology.

Proof: Because of Theorem 3.1 and the fact that V'_σ maps $\mathcal{H}^{(n)} \cap D((N_a+1)^{1/2})$ into $\mathcal{H}^{(n)} \oplus \mathcal{H}^{(n-1)} \oplus \mathcal{H}^{(n+1)}$ one can carry over the method that Lanford used for the proof of the corresponding statement in the cut-off Yukawa theory (Ref. 41, Chap. III, Sec. 2, Proposition 3.4., p. 74). ∎

Now we come back to the proof of point (i) of Theorem 3.2b. We have proved by now that $B^*(e^{-it\omega}h'_\sigma)e^{-it\hat{H}_\sigma} \in \mathcal{L}(D^{(n);\alpha}; D^{(n+1);\alpha'})$, for all $\alpha > (1/2)\ln 2, 0 < \alpha' < \alpha - (1/2)\ln 2$. Using again Lemma 3.2, we conclude $\hat{b}^{\#}_{\sigma,t}(h'_\sigma) \in \mathcal{L}(D^{(n);\alpha}; D^{(n\pm1);\alpha'})$.

(ii) All commutators (3.19a) are defined on $D''^{(n)}$ by the estimates (i) on the domain and ranges of the operators involved. The O-commutation relations are then immediate consequences of the definitions. To prove the non-O-commutation relation (3.19b) it suffices to show that

$$[B(f'_\sigma), B^*(g'_\sigma)]\Omega_0 = (f,g).$$

But on Ω_0 the commutator reduces to (cf. Remark 3.3)

$$[b(\bar{f}'_\sigma), b^*(g'_\sigma)]\Omega_0 + b(\bar{f}'_\sigma)\underbrace{\Gamma(Q_\sigma)^*\Gamma(Q_\sigma)}_{1}b^*(g'_\sigma)\Omega_0 + b(\bar{f}'_\sigma)\Gamma(Q_\sigma)\underbrace{b^*(g'_\sigma)\Omega_0}_{1}.$$

Taking the scalar product with Ω_0 we obtain (realizing that some terms vanish)

$$\begin{aligned}(\Omega_0, [B(\bar{f}'_\sigma), B^*(g'_\sigma)]\Omega_0) &= (b^*(\bar{f}'_\sigma)\Omega_0, \{1 + \Gamma(Q_\sigma)^*\Gamma(Q_\sigma)\}b^*(g'_\sigma)\Omega_0) \\ &= (b^*(\bar{f}'_\sigma)\Omega_0, A_\sigma^{-2}b^*(g'_\sigma)\Omega_0) = (b^*(\bar{f})\Omega_0, b^*(g)\Omega_0) \\ &= (f,g),\end{aligned}$$

where the fact that: $[\Gamma(Q_\sigma)^*\Gamma(Q_\sigma)]\restriction_M : b^*(g'_\sigma)\Omega_0 = 0$ has been used, together with the definition of A_σ.

(iii) For $\sigma < \infty$ we have only to use that $e^{it\hat{H}_\sigma} = e^{itH_0}$ on $\mathcal{H}^{(0)}$ and $e^{-it\hat{H}_\sigma}T_\sigma A_\sigma b^*(h)\Omega_0 = T_\sigma A_\sigma b^*(e^{-it\omega}h)\Omega_0$, as shown using Theorem 3.1, first for $h \in \mathfrak{D}$ and then for all h (using the fact that $e^{-it\hat{H}_\sigma}T_\sigma A_\sigma$ is bounded on $\mathcal{H}^{(1,0)}$). For $\sigma = \infty$ we use that on $\mathcal{H}^{(0)} \cup \mathcal{H}^{(1,0)}_{\text{phys}}(R)$ one has $e^{it\hat{H}'_\infty} = \operatorname{s-lim}_{\sigma\to\infty} e^{it\hat{H}'_{\sigma(R)}} = e^{itH_0}$, together with the Theorem 3.1b and Lemma 3.1b.

(iv) Using Lemma 3.2 and the first order estimate of Lemma 3.1a, we have that, for $\sigma < \infty$, $e^{-it\hat{H}_\sigma}$ maps $D(H_0^{1/2}) \cap D^{(n);\alpha}$ into $D(H_0^{1/2}) \cap D^{(n);\alpha''}$, for any $0 < \alpha'' < \alpha$. Using then $a^{\#}(g) \in \mathcal{L}_{\alpha''\alpha'}$, for any $0 < \alpha' < \alpha''$, and again Lemma 3.2, we conclude $\hat{a}^{\#}_{\sigma,t}(g) \in \mathcal{L}_{\alpha\alpha'}$. The remaining part is shown observing that all the assumptions used in Sec. 2 are satisfied.

(v) This follows easily from the already proved properties on domains and ranges of the $\hat{b}^{\#}, \hat{a}^{\#}$ operators. This concludes the proof of Theorem 3.2b. ∎

We will now give expansions of the dressed fields in power series in the coupling constant.

Let η be a label taking the values a (for meson) and b (for nucleons). Set, for any $h^{(\eta)} \in L_2(\mathbb{R}^3)$, $\hat{h}^{(\eta)}_\sigma \equiv (\nu^{(\eta)}_\sigma)^{-1/2}h^{(\eta)}$, with $\nu^{(b)} \equiv \nu^{-1/2}_\sigma$, $\nu^{(a)}_\sigma \equiv 1$.

Define then

$$\begin{aligned}\hat{c}^{(\eta)\#}_{\sigma,t}(\hat{h}^{(\eta)}_\sigma) &\equiv \begin{cases}\hat{b}^{\#}_{\sigma,t}(\nu^{-1/2}_\sigma h^{(b)}) & \text{for } \eta = b\\ \hat{a}^{\#}_{\sigma,t}(h^{(a)}) & \text{for } \eta = a\end{cases} \\ &= e^{it\hat{H}_\sigma}C^{(\eta)\#}(e^{\mp it\Omega^{(\eta)}}\hat{h}^{(\eta)}_\sigma)e^{-it\hat{H}_\sigma},\end{aligned}$$

with

$$C^{(\eta)\#}(\cdot) \equiv \begin{cases}B^{\#}(\cdot) & \text{for } \eta = b\\ a^{\#}(\cdot) & \text{for } \eta = a\end{cases},$$

$$\Omega^{(\eta)}(\cdot) \equiv \begin{cases}\omega(\cdot) & \text{for } \eta = b\\ \mu(\cdot) & \text{for } \eta = a\end{cases}.$$

Thus $\hat{c}^{(\eta)\#}_{\sigma,t}(\cdot)$ are the time t Heisenberg picture's adjusted creators and annihilators for dressed b and a particles.

We have:

Theorem 3.3: For any $0 \le \sigma < \infty$ there exist numbers $\lambda^{(\eta)}(\sigma) > 0$ such that for all $|\lambda| \le \lambda^{(\eta)}(\sigma)$ the following expansions of the fields in powers of λ hold for all $\Phi \in \overset{0}{D}{}^{(n)}$:

$$\hat{c}^{(\eta)\#}_{\sigma,t}(\hat{h}^{(\eta)}_\sigma)\Phi = \sum_{l=0}^{\infty} C^{(l)(\eta)\#}_{\sigma,t}(\hat{h}^{(\eta)}_\sigma)\Phi, \tag{3.25}$$

where the series on the right-hand side converges strongly and

$$\begin{aligned}C^{(l)(\eta)\#}_{\sigma,t}(\hat{h}^{(\eta)}) \equiv (i)\int\cdots\int &[V'(t_l), \ldots, [V'(t_1), \\ &e^{itH_0}C^{(\eta)\#}_\sigma(e^{\mp it\Omega^{(\eta)}}\hat{h}^{(\eta)}_\sigma)e^{-itH_0}]\cdots]dt_1\cdots dt_l\ , \\ &\times 0 \le t_l \le \ldots \le t_1 \le t\end{aligned} \tag{3.26}$$

with

$$V'_\sigma(s) \equiv e^{isH_0}V'_\sigma e^{-isH_0}, \qquad V'_\sigma = \lambda V_\sigma + \hat{M}_\sigma.$$

V'_σ itself is a power series in λ and in (3.25) one can replace $C^{(l)(\eta)\#}_{\sigma,t}(\hat{h}^{(\eta)}_\sigma)$ by the power series one obtains expanding each $V'_\sigma(\cdot)$ occurring in the definition (3.26) by its correspondent power series in λ. Moreover, also $C^{(\eta)\#}_\sigma(\cdot)$ and $\hat{h}^{(\eta)}_\sigma$ can be expanded in power series of λ. Each of these additional expansions conserves the strong convergence of the series. Hence $\hat{c}^{(\eta)\#}_{\sigma,t}(\hat{h}^{(\eta)}_\sigma)$ are for all $|\lambda| \le \lambda^{(\eta)}(\sigma)$, all t, holomorphic in λ.

Proof: Use of $\overset{0}{D}{}^{(n)} \subset \cup_{\alpha>0}D^{(n);\alpha}$, Lemma 3.1b, Theorems 3.2a, 3.2b give expansions like the one in Lemma 3.2, convergent in the $\mathcal{L}_{\alpha\beta}$-topology, $0 < \beta < \alpha$. Then the proof follows by reordering of the terms. ∎

Remark 3.4: For $\eta = a$ (meson fields) we have the simple power series expansions

$$\hat{a}^{\#}_{\sigma,t}(f) = \sum_{l=0}^{\infty}\lambda^l I^{(l)}_\sigma(f;t), \tag{3.27}$$

where the $I^{(l)}_\sigma(f;t)$ are independent of λ, $I^{(0)}_\sigma(f;t) \equiv a^*(h)$ and $\lambda^l I^{(l)}_\sigma(f;t)$ for $l = 1, 2, \cdots$ is the sum of all terms of order l in λ one can extract from

$$(i)^l \int_{0\le t_l\le\ldots\le t_1\le t} [V'_\sigma(t_l), [V'_\sigma(t_{l-1}),\ldots,[V'_\sigma(t_1),a^{\#}(h)]\cdots]]dt_1\ldots dt_l. \quad (3.28)$$

This expansion is of the same type as those considered by Höegh-Krohn in Ref. 23b.

The wave operators. Strong asymptotic convergence of states

We introduced already the operators $\Omega_\sigma(t)$. We shall prove that they converge as $t \to \pm\infty$ to wave operators. Such operators have been introduced in this model by Eckmann.[20a]

Let $D^{\beta(n)}_{(\sigma)} \equiv D^{\beta(n)}$ for $\sigma < \infty$ and $D^{\beta(n)}_{(\sigma)} \equiv D^{\beta(n)} \cap \mathcal{H}^{(n)}(R)$ for $\sigma = \infty$, where $D^{\beta(n)} \equiv \cup_{\alpha>(\beta/2)\ln 2} D^{(n);\alpha}$.

We have:

Theorem 3.4: For any $0 \le \sigma < \infty$ there exists a number $\Lambda_0(\sigma) > 0$, and for $\sigma = \infty$ and any given $0 < R < \infty$ there exists a number $\Lambda_0(R) > 0$ such that for all $|\lambda| \le \Lambda$, with $\Lambda = \Lambda_0(\sigma)$ for $\sigma < \infty$ and $\Lambda = \Lambda_0(R)$ for $\sigma = \infty$, one has:

(i) The operators $\Omega_\sigma(t) \equiv e^{it\hat{H}}{}_\sigma T_\sigma A_\sigma e^{-itH_0}$ are defined on the dense subset $D'^{(n)}_{(\sigma)}$ of $\mathcal{H}^{(n)}_{(\sigma)}$ and

$$\Omega_\sigma(t) \in \mathcal{L}(D^{(n);\alpha}_{(\sigma)}; D^{(n);\alpha'}) \subset \mathcal{L}(D^{(n);\alpha}_{(\sigma)}; \mathcal{H}^{(n)})$$

for any $0 < \alpha' < \alpha$, with relative operator norms bounded uniformly in time:

$$\|\Omega_\sigma(t)\|_{(D^{(n);\alpha}_{(\sigma)};\mathcal{H}^{(n)})} \le C_1 \|\Omega_\sigma(t)\|_{\alpha\alpha'} < C_2, \quad (3.29)$$

with constants C_1, C_2 independent of t.

(ii) The strong limits for $t \to \pm\infty$ of $\Omega_\sigma(t)$ with respect to the topology of $\mathcal{H}^{(n)}$ exist on $D'^{(n)}_{(\sigma)}$:

$$\text{s-}\lim_{t\to\pm\infty} \Omega_\sigma(t)\Psi = \Omega^\pm_\sigma\Psi, \quad (3.30)$$

for all $\Psi \in D'^{(n)}_{(\sigma)}$.

The generalized wave operators $\Omega^\pm_\sigma$ are defined on the dense set $D'^{(n)}_{(\sigma)}$ of $\mathcal{H}^{(n)}_{(\sigma)}$ and belong to $\mathcal{L}(D^{(n);\alpha}_{(\sigma)}; \mathcal{H}^{(n)})$ for any $\alpha > \frac{1}{2}\ln 2$.

(iii) For all $\Psi \in D''^{(n)}_{(\sigma)} \equiv D^{2(n)}_{(\sigma)}$ one has

$$(\Omega^\pm_\sigma)^*\Omega^\pm_\sigma\Psi = \Psi \quad (3.31)$$

and

$$e^{it\hat{H}}{}_\infty \Omega^\pm_\infty\Psi = \Omega^\pm_\infty e^{itH_0}\Psi. \quad (3.32)$$

(iv) $(\Omega^\pm_\sigma)^*$ are partial isometries with initial domains the closures $R^\pm_\sigma$ of $\Omega^\pm_\sigma D''^{(n)}_{(\sigma)}$ in $\mathcal{H}^{(n)}_{(\sigma)}$ and final domains $\mathcal{H}^{(n)}$. $\Omega^\pm_\sigma$ can be extended (from $D''^{(n)}_{(\sigma)}$) in a unique way to partial isometries $\tilde{\Omega}^\pm_\sigma$ with initial domain $\mathcal{H}^{(n)}$ and final domains $\Omega^\pm_\sigma D''^{(n)}_{(\sigma)} = \tilde{\Omega}^\pm_\sigma \mathcal{H}^{(n)}_{(\sigma)}$. Note $(\tilde{\Omega}^\pm_\sigma)^* = (\Omega^\pm_\sigma)^*$. The relation (3.31) is then extended to $\tilde{\Omega}^{\pm *}_\sigma\tilde{\Omega}^\pm_\sigma = 1$ on $\mathcal{H}_{(\sigma)} = \oplus_n \mathcal{H}^{(n)}_{(\sigma)}$ and the relations (3.32) to $e^{it\hat{H}}{}_\sigma\tilde{\Omega}^\pm_\sigma = \tilde{\Omega}^\pm_\sigma e^{itH_0}$ on $\mathcal{H}_{(\sigma)}$.

Moreover,

$$\tilde{\Omega}^\pm_\sigma(\tilde{\Omega}^\pm_\sigma)^* = P_{R^\pm_\sigma}, \quad (3.33)$$

where $P_{R^\pm_\sigma}$ are the projectors on R^+_σ, R^-_σ, respectively.[43,44]

Remark 3.5: On $\mathcal{H}^{(0)} \cup \mathcal{H}^{(1,0)}$ one has $\tilde{\Omega}^\pm_\sigma = \Omega^\pm_\sigma = \Omega_\sigma(t) = T_\sigma A_\sigma$. Hence, in particular, $\tilde{\Omega}^\pm_\sigma$ map $\mathcal{H}^{(1,0)}$ (all bare one nucleon states) isometrically into $\mathcal{H}^{(1,0)}_{(\sigma)\text{phys}} \equiv T_\sigma A_\sigma \mathcal{H}^{(1,0)}$ (all physical one nucleon states). Moreover, $\tilde{\Omega}^\pm_\sigma = 1 = (\tilde{\Omega}^\pm_\sigma)^*$ on $\mathcal{H}^{(0)}$.

Proof of Theorem 3.4: The proof for $\sigma = \infty$ parallels completely the one for $\sigma < \infty$. Let us write the details only for $\sigma < \infty$.

(i) Using Lemma 3.2, the fact that $T_\sigma \in \mathcal{L}(D^{(n);\beta}; D^{(n);\beta'})$ for any $0 < \beta' < \beta - (\frac{1}{2})\ln 2$ (Lemma 3.1a) and the fact that $A_\sigma e^{itH_0}$ leaves each $D^{(n);\alpha}$ invariant, we have immediately $\Omega_\sigma(t) \in \mathcal{L}(D^{(n);\alpha}; D^{(n);\alpha'})$, $0 < \alpha' < \alpha - (\frac{1}{2})\ln 2$. The stated uniform bounds follow from the isometry of e^{itH_0} on each $D^{(n);\alpha}$, in the $D^{(n);\alpha}$-topology.

(ii) We first remark that, due to the uniform bound on $\Omega_\sigma(t)$ in $\mathcal{L}_{\alpha\alpha'}$, for all t, it is sufficient to prove (3.30) on some subset of $D^{(n);\alpha}$, dense in $D^{(n);\alpha}$ in the $\|\ \|_\alpha$-topology (any $\alpha > (\frac{1}{2})\ln 2$).

This will be done in the following taking $D = \mathring{D}^{(n)}, \mathring{D}^{(n)}$ as in Remark 3.1. First we prove the following:

Lemma 3.3: Let σ, λ be as in Theorem 3.4, $\Phi \in \mathring{D}^{(n)}$. Then $\Omega_\sigma(t)\Phi$ is strongly differentiable with respect to t and, in the sense of strong derivatives,

$$\frac{d\Omega_\sigma}{dt}(t)\Phi = ie^{+it\hat{H}}{}_\sigma(\hat{H}_\sigma T_\sigma A_\sigma - T_\sigma A_\sigma H_0)e^{-itH_0}\Phi$$
$$= ie^{+it\hat{H}}{}_\sigma[:T_\sigma\{\lambda V^a_\sigma + \lambda V^a_\sigma \underbrace{\Gamma(Q_\sigma)}_{1} + \tfrac{1}{2}\lambda V^a_\sigma \underbrace{:(\Gamma(Q_\sigma))^2:}_{2}\}]A_\sigma e^{-itH_0}\Phi. \quad (3.34a)$$

Moreover, the strong derivative $d\Omega_\sigma/dt$ is strongly continuous in t and

$$\Omega_\sigma(t)\Phi = \Omega_\sigma(0)\Phi + i\int_0^t dt' e^{+it'\hat{H}}{}_\sigma[:T_\sigma\{\lambda V^a_\sigma + \lambda V^a_\sigma \underbrace{\Gamma(Q_\sigma)}_{1} + \frac{\lambda}{2} V^a_\sigma \underbrace{:(\Gamma(Q_\sigma))^2:}_{2}\}:]A_\sigma e^{-it'H_0}\Phi, \quad (3.34b)$$

the integral being understood in the strong sense.

Proof of Lemma 3.3: Let $U(t) \equiv e^{it\hat{H}}{}_\sigma$, $U_0(t) \equiv e^{-itH_0}$. Since (using Lemma 3.1 and the strong differentiability of $U_0(t)$ in both the $\mathcal{H}^{(n)}$ and $\|\ \|_\alpha$-topologies)

$$\lim_{\epsilon\to 0} \epsilon^{-1}[U(t)T_\sigma A_\sigma\{U_0(t+\epsilon) - U_0(t)\}\Phi] = -iU(t)T_\sigma A_\sigma H_0 U_0(t)\Phi,$$

for (3.34a) it suffices to prove

$$\lim_{\epsilon\to 0} \epsilon^{-1}[U(t+\epsilon)T_\sigma A_\sigma U_0(t+\epsilon)\Phi - U(t)T_\sigma A_\sigma U_0(t+\epsilon)\Phi] = U(t)\hat{H}_\sigma T_\sigma A_\sigma U_0(t)\Phi. \quad (3.35)$$

Using the fact that the vectors of $\mathring{D}^{(n)}$ are entire vectors of H_0, together with Lemma 3.1, we easily obtain (3.35). The strong continuity in t of $(d\Omega_\sigma/dt)\Phi$ follows from the strong continuity of $U(t)$ in the $\mathcal{H}^{(n)}$-topology, together with Lemma 3.1, Remark 3.1, and the strong continuity of $A_\sigma U_0(t)$ in the $\|\ \|_\alpha$-topology. ■

To finish the proof of point (ii) of Theorem 3.4 it suffices now to estimate

$$\int_0^t \left\| \frac{d\Omega_\sigma}{dt'}(t')\Phi \right\| dt',$$

for $t \to \pm\infty$, Φ in some dense subset of $\mathring{D}^{(n)}$. Using the representation of $(d\Omega_\sigma/dt')(t')$ given by the right-hand side of the second equality in (3.34a), we see that the bracket $\{\ \}$ on the right-hand side is a sum of connected terms (in the terminology of, e.g., K. Hepp, Ref. 21, p. 16) containing two uncontracted annihilation operators. Thus the terms in $\{\ \}$ have the same "operator form" as the ones in K. Hepp, Ref. 21 [Theorem 3.3, formula (3.39)].

We can carry over the arguments of Hepp; Lemma 10 of Ref. 20a allows us to expand the bracket $[\ ]$ into a sum of terms of different order of Q_σ, T_σ. Call $\mathcal{T}_\nu^{(i)}$ a term of order ν with i meson creators occurring in this expansion.

It suffices to estimate $\|\mathcal{T}_\nu^{(i)} A_\sigma e^{-itH_0}\Phi\|^2$ for all $\Phi \in \mathring{\mathfrak{N}}^{(n)}$, where $\mathring{\mathfrak{N}}^{(n)}$ is the following subset of $\mathring{D}^{(n)}$, consisting of functions with "nonoverlapping" velocities:

$$\mathring{\mathfrak{N}}^{(n)} \equiv \{\Psi = \{\Psi^{(n,m)}; m = 0,1,2,\dots\} \in \mathring{D}^{(n)} \mid \times \Psi^{(n,m)}(q_1 \cdots q_n; p_1 \cdots p_m) \text{ such that if } p_i, q_j \times \in \operatorname{supp}\Psi^{(n,m)}, \text{then } p_i/\mu_i \neq p_k/\mu_k \text{ for all } i \neq k, \quad q_j/\omega_j \neq q_l/\omega_l \text{ for all } j \neq l, p_i/\mu_i \neq q_j/\omega_j \text{ for all } i,j\}.$$

One has, performing two partial integrations:

$$\|\mathcal{T}_\nu^{(i)} A_\sigma e^{-itH_0}\Phi\|^2 \leq C_\nu^{(i)}(\Phi)(1+t^2)^{-1},$$

where $C_\nu^{(i)}$ is independent of t (depends on Φ and its partial derivatives up to order 2). By a computation similar to Eckmann's one, which lead to an estimate of $\|T_\sigma \chi\|$, $\chi \in D^{(n),\alpha}$ [Lemma 3.1b, (i)], one can show $\sum_{\nu,i} C_\nu^{(i)}(\Phi) < \infty$. This has a twofold reason. On one hand one can estimate the single terms $C_\nu^{(i)}(\Phi)$ by "kernel norms" (in the terminology of Ref. 20a), controlled in the same way as those of the terms of order ν and with i meson creators out of $:T_\sigma \lambda V_\sigma^a:, :T_\sigma(\lambda V_\sigma^a \underset{1}{\smile} \Gamma(Q_\sigma):, \tfrac{1}{2}:T_\sigma(\lambda V_\sigma^a \underset{2}{\smile} (\Gamma(Q_\sigma))^2):$, thus obtaining $C_\nu^{(i)} \leq C_1 C_2^\nu (i!)^{-1+\epsilon}$, $\epsilon > 0$ (C_1, C_2 being constants), with $\sum_\nu C_2^\nu < \infty$ (the smallness of $|\lambda|$ being used at this place). On the other hand the number of terms $C_\nu^{(i)}$ for given ν is only superior to the number of terms of order ν in T_σ by a factor increasing at most polynomially in ν.

But $\sum_\nu \sum_i C_\nu^{(i)}(\Phi) \leq C_1' \sum_\nu \nu^k C_2^\nu < \infty$ (for some $k > 0$) and this concludes then the proof of the strong convergence of $\Omega_\sigma(t)$. The property (3.32) follows then from

$$e^{it\hat{H}_\sigma}\Omega_\sigma(s) = \Omega_\sigma(t+s)e^{itH_0}$$

on $D'^{(n)}$, taking the limit for $s \to \pm\infty$. The property (3.31) follows by computing, for $\Psi \in D''^{(n)}$, $\varphi \in D'^{(n)}$:

$$\lim_{t\to\pm\infty} (\varphi, e^{itH_0} A_\sigma T_\sigma^* T_\sigma A_\sigma e^{-itH_0}\Psi). \quad \blacksquare$$

We shall now study the convergence to asymptotic fields. The following theorem is an immediate consequence of the results in Sec. 2 and the preceding theorems:

Theorem 3.5:

(i) For given $0 \leq \sigma < \infty$ or $\sigma = \infty$, $0 < R < \infty$ let λ be such that $\hat{a}_{\sigma,t}^{\#}(h)$ exists [Theorem 3.2b, (iv)]. Then the conclusions of Theorem 2.2, 2.3 hold for all $\sigma \leq \infty$ and yield the asymptotic meson fields $\hat{a}_{\infty,\pm}^{\#}(h)$. These have the free commutation relations among themselves and the same commutation relations with $e^{it\hat{H}_\infty}$ as the free fields with e^{itH_0}.

On $\mathcal{K}^{(0)}$ one has $\hat{a}_{\infty,\pm}^*(h) = \hat{a}_{\sigma,t}^*(h) - a^*(h)$ and on $\mathcal{K}^{(0)} \cup \mathcal{K}^{(1,0)}$: $\hat{a}_{\infty,\pm}^*(h) T_\infty A_\infty = \hat{a}_{\infty,\pm}^*(h)\Omega_\infty^\pm = \Omega_\infty^\pm a^*(h)$.

(ii) For given $0 \leq \sigma < \infty$ or $\sigma = \infty$, $0 < R < \infty$ there exist $\Lambda_2(\sigma) > 0$ resp. $\Lambda_2(R) > 0$ such that for all $\sigma \leq \infty$ and all $|\lambda| \leq \Lambda_2$ and all $h \in L_2(\mathbb{R}^3)$: s-lim $b_{\sigma,t}^*(h_\sigma')\Omega_\sigma^\pm$ exist on $D'^{(0)} \cup \mathcal{K}_{(\sigma)}^{(1,0)}$ and satisfy

$$\text{s-lim}_{t\to\pm\infty} \hat{b}_{\sigma,t}^*(h_\sigma')\Omega_\sigma^\pm = \hat{b}_{\sigma,\pm}^*(h_\sigma')\Omega_\sigma^\pm = \Omega_\sigma^\pm b^*(h).$$

In particular, $\hat{b}_{\sigma,\pm}^*(h_\sigma')\Omega_0 = T_\sigma A_\sigma b^*(h)\Omega_0$.

Moreover, $\hat{b}_{\sigma,\pm}(h_\sigma')\phi = \text{s-lim}_{t\to\pm\infty} \hat{b}_{\sigma,t}(h_\sigma')\phi = 0$ for all $\phi \in \mathcal{K}^{(0)}$.

(iii) For λ, h as in (ii) and $f \in L_2(\mathbb{R}^3)$ one has

$$\text{s-lim}_{t\to\pm\infty} \hat{b}_{\sigma,t}^*(h_\sigma')\hat{a}_{\sigma,t}^*(f)\Omega_0 = \hat{b}_{\sigma,\pm}^*(h_\sigma')\hat{a}_{\sigma,\pm}^*(f)\Omega_0 = \hat{a}_{\sigma,\pm}^*(f)\hat{b}_{\sigma,\pm}^*(h_\sigma')\Omega_0 = \hat{a}_{\sigma,\pm}^*(f)B^*(h_\sigma')\Omega_0. \quad \blacksquare$$

Remark 3.6: The statements (ii), (iii) give partial information on the asymptotic nucleon fields. These results will be generalized to the strong convergence of the dressed nucleon fields on states with arbitrarily many nucleons and mesons in the second part of this series of papers.[26]

B. Some connections between asymptotic states and fields. The S matrix

Theorem 3.6: For any $0 \leq \sigma < \infty$ and, if $\sigma = \infty$, for any $0 < R < \infty$, there exist numbers $\Lambda_0(\sigma) > 0$ resp. $\Lambda_0(R) > 0$ such that, for all $|\lambda| \leq \Lambda_0$:

(1) for all $h \in L_2(\mathbb{R}^3)$ and all $\Phi \in D_{(\sigma)}^{\prime(n)}$ one has

$$\text{s-lim}_{t\to\pm\infty} \hat{a}_{\sigma,t}^*(h)\Omega_\sigma(t)\Phi = \hat{a}_{\sigma,\pm}^*(h)\Omega_\sigma^\pm\Phi.$$

Moreover,

$$\hat{a}_{\sigma,\pm}^*(h)\Omega_\sigma^\pm\Phi = \Omega_\sigma^\pm a^*(h)\Phi,$$

i.e.,

$$(\tilde{\Omega}_\sigma^\pm)^*\hat{a}_{\sigma,\pm}^*\tilde{\Omega}_\sigma^\pm = a^*(h), \text{ on } D_{(\sigma)}^{\prime(n)}.$$

(2) The product $\prod_{i=1}^m \hat{a}_{\sigma,t}^*(h^{(i)})\Psi$ is defined for all $h^{(i)} \in L_2(\mathbb{R}^3)$ and all $\Psi \in D^{\prime(0)} \cup \mathcal{K}_{(\sigma)\text{phys}}^{(1,0)}$, where $\mathcal{K}_{(\sigma)\text{phys}}^{(1,0)}$ is equal $T_\sigma A_\sigma \mathcal{K}^{(1,0)}$.

One has

$$\text{s-lim}_{t\to\pm\infty} \prod_i \hat{a}_{\sigma,t}^*(h^{(i)})\Psi = \prod_i \hat{a}_{\sigma,\pm}^*(h^{(i)})\Psi = \Omega_\sigma^\pm \prod_i a^*(h^{(i)})\Psi.$$

(3) One has, for all $\Psi \in \mathcal{K}_{(\sigma)}^{(1,0)}$,

$$\prod_{j=1}^n \hat{b}_{\sigma,t}^*(h_\sigma'^{(j)}) \prod_{i=1}^m \hat{a}_{\sigma,t}^*(f^{(i)})\Psi = \Omega_\sigma(t) \prod_j b^*(h^{(j)}) \prod_i a^*(f^{(i)})\Psi$$

and both sides converge strongly to

$$\Omega_\sigma^\pm \prod_j b^*(h^{(j)}) \prod_i a^*(f^{(i)})\Psi.$$

(4) For all $\Phi \in D^{\prime(0)}$, $h \in \mathcal{K}_{(\sigma)}^{(1,0)}$, $f^{(i)} \in L_2(\mathbb{R}^3)$:

$$\text{s-lim}_{t\to\pm\infty} \hat{b}_{\sigma,t}^*(h_\sigma') \prod_{i=1}^m \hat{a}_{\sigma,t}^*(f^{(i)})\Phi = \hat{b}_{\sigma,\pm}^*(h_\sigma') \prod_{i=1}^m \hat{a}_{\sigma,\pm}^*(f^{(i)})\Phi = \Omega_\sigma^\pm b^*(h) \prod_i a^*(f^{(i)})\Phi.$$

(5) Call $\theta_{(\sigma),\pm}^{(n)}$ the closed linear span of all states of the form

$$\text{s-lim}_{t\to\pm\infty} \prod_{j=1}^{n} \hat{b}^*_{\sigma,t}(g'^{(j)}_\sigma) \prod_{i=1}^{m} \hat{a}^*_{\sigma,t}(f^{(i)})\Omega_0,$$

$$g^{(j)} \in \mathcal{K}^{(1,0)}_{(\sigma)}, f^{(i)} \in L_2(\mathbb{R}^3),$$

for n fixed, $m = 0, 1, 2, \cdots$, with convention $\Pi^m \equiv 1$ for $m = 0$. $\theta^{(n)}_{(\sigma),\pm}$ is a subspace of $\mathcal{K}^{(n)}_{(\sigma)}$. $\mathcal{K}^{(n)}_{(\sigma)}$ is mapped by $\tilde{\Omega}^{\pm}_\sigma$ into $\theta^{(n)}_{(\sigma),\pm}$. The scattering operator $\hat{S}_\sigma$ is defined by $\hat{S}_\sigma = (\tilde{\Omega}^+_\sigma)^* \tilde{\Omega}^-_\sigma$ as a contraction from $\mathcal{K}^{(n)}_{(\sigma)}$ into $\mathcal{K}^{(n)}_{(\sigma)}$.[45] $\hat{S}_\sigma$ extends trivially to a contraction in $\mathcal{K}_{(\sigma)} \equiv \oplus_{n=0}^{\infty} \mathcal{K}^{(n)}_{(\sigma)}$.

(6) The amplitude for the scattering from an incoming state

$$\text{s-lim}_{t\to-\infty} \prod_{j=1}^{n} \hat{b}^*_{\sigma,t}(g'^{(j)}_\sigma) \prod_{i=1}^{m} \hat{a}^*_{\sigma,t}(f^{(i)})\Omega_0$$

to an outgoing state

$$\text{s-lim}_{t\to+\infty} \prod_{j=1}^{n} \hat{b}^*_{\sigma,t}(\tilde{g}'^{(j)}_\sigma) \prod_{i=1}^{m'} \hat{a}^*_{\sigma,t}(\tilde{f}^{(i)})\Omega_0,$$

$$\tilde{g}^{(j)} \in \mathcal{K}^{(1,0)}_{(\sigma)}, \tilde{f}^{(i)} \in L_2(\mathbb{R}^3),$$ is given by

$$\Big(\text{s-lim}_{t\to+\infty} \prod_j \hat{b}^*_{\sigma,t}(\tilde{g}'^{(j)}_{(\sigma)}) \prod_{i=1}^{m'} \hat{a}^*_{\sigma,t}(\tilde{f}^{(i)})\Omega_0,$$

$$\text{s-lim}_{t\to-\infty} \prod_j \hat{b}^*_{\sigma,t}(g'^{(j)}_\sigma) \prod_{i=1}^{m} \hat{a}^*_{\sigma,t}(f^{(i)})\Omega_0\Big)$$

$$= \Big(\Omega^+_\sigma \prod_j b^*(\tilde{g}^{(j)}) \prod_{i=1}^{m'} a^*(\tilde{f}^{(i)})\Omega_0,$$

$$\Omega^- \prod_j b^*(g^{(j)}) \prod_{i=1}^{m} a^*(f^{(i)})\Omega_0\Big)$$

$$= \Big(\prod_j b^*(\tilde{g}^{(j)}) \prod_{i=1}^{m'} a^*(\tilde{f}^{(i)})\Omega_0,$$

$$\hat{S}_\sigma \prod_j b^*(g^{(j)}) \prod_{i=1}^{m} a^*(f^{(i)})\Omega_0\Big)$$

For $n = 1$ we have also

$$\Big(b^*(\tilde{g}) \prod_{i=1}^{m'} a^*(\tilde{f}^{(i)})\Omega_0, \hat{S}_\sigma b^*(g) \prod_{i=1}^{m} a^*(f^{(i)})\Omega_0\Big)$$

$$= \Big(\hat{b}^*_{\sigma,-}(\tilde{g}'_\sigma) \prod_{i=1}^{m'} \hat{a}^*_{\sigma,-}(\tilde{f}^{(i)})\Omega_0, \hat{b}^*_{\sigma,+}(g'_\sigma) \prod_{i=1}^{m} \hat{a}^*_{\sigma,+}(f^{(i)})\Omega_0\Big)$$

Proof: All points are proved using essentially the uniform bound in t on $\Omega_\sigma(t)$ resp. $\Omega_\infty(t)$ as an operator from $D^{(n);\alpha}$, $\alpha > (\frac{1}{2}) \ln 2$ to $\mathcal{K}^{(n)}$ (Theorem 3.4) as well as Theorem 3.5 and the equality $e^{it\hat{H}_\sigma} = e^{itH_0}$ on $\mathcal{K}^{(0)} \cup \mathcal{K}^{(1,0)}_{(\sigma)\text{phys}}$. ■

Remark 3.9: Above results will be generalized in part II of this series of papers (Ref. 26).

Note that, of course, $\hat{S}_\sigma = 1$ on $\mathcal{K}^{(0)}$.

C. Reduction formulas for the meson-nucleon scattering

Let us consider the scattering of one physical nucleon and one meson. We shall give some remarks for both cases $\sigma < \infty$ and $\sigma = \infty$, using again the notation of the preceding section.

Let the asymptotic situation at $t = -\infty$ be described by

$$|h^{(1)}; f^{(1)}\rangle_- \equiv \hat{b}^*_{\sigma,-}(h'^{(1)}_\sigma)\hat{a}^*_{\sigma,-}(f^{(1)})\Omega_0$$

and the asymptotic situation at $t = +\infty$ be described by

$$|h^{(2)}; f^{(2)}\rangle_+ \equiv \hat{b}^*_{\sigma,+}(h'^{(2)}_\sigma)\hat{a}^*_{\sigma,+}(f^{(2)})\Omega_0,$$

$$h^{(i)} \in \mathcal{K}^{(1,0)}_{(\sigma)}, f^{(i)} \in L_2(\mathbb{R}^3), i = 1, 2.$$

The S matrix element for the transition is, according to Theorem 3.6,

$$\hat{S}_\sigma(h^{(1)}, f^{(1)}; h^{(2)}, f^{(2)}) = (\hat{b}^*_{\sigma,-}(h'^{(1)}_\sigma)\hat{a}^*_{\sigma,-}(f^{(1)})\Omega_0,$$

$$\hat{b}^*_{\sigma,+}(h'^{(2)}_\sigma)\hat{a}^*_{\sigma,+}(f^{(2)})\Omega_0)$$

$$= (b^*(h^{(1)})a^*(f^{(1)})\Omega_0, \hat{S}_\sigma b^*(h^{(2)})a^*(f^{(2)})\Omega_0) \qquad (3.36)$$

We shall first derive simple reduction formulas which are related to the well known ones of the relativistic formalism and shall be the starting point for our preliminary discussion of the meson-nucleon scattering in the model.

We first observe using Theorem 3.5

$$\hat{b}^*_{\sigma,\pm}(h'_\sigma)\hat{a}^*_{\sigma,\pm}(f)\Omega_0 = \hat{a}^*_{\sigma,\pm}(f)\hat{b}^*_{\sigma,\pm}(h'_\sigma)\Omega_0$$

$$= \hat{a}^*_{\sigma,\pm}(f)B^*(h'_\sigma)\Omega_0, \quad \text{for any } f, h \in L_2(\mathbb{R}^3).$$

Using furthermore (Theorems 2.3, 3.5),

$$\hat{a}^*_{\sigma,\pm}(f) = a^*(f) + i\int_0^{\pm\infty} \frac{d}{dt'} \hat{a}^*_{\sigma,t'}(f)dt'$$

$$= a^*(f) + i\int_0^{\pm\infty} e^{it\hat{H}_\sigma} \mathfrak{A}^*_\sigma(f;t) e^{-it\hat{H}_\sigma} dt, \qquad (3.37)$$

where f stands for $f^{(1)}$ and $f^{(2)}$, $\mathfrak{A}^*_\sigma(f;t) \equiv \mathfrak{A}^*_\sigma(f_t)$ for the bounded extension to $\mathcal{K}^{(n)}$ of the bounded, densely defined operator $\lambda[V_\sigma, a^*(e^{-it\mu}f)]$ (as in Lemma 2.2), we have

$$\hat{S}_\sigma(h^{(1)}, f^{(1)}; h^{(2)}, f^{(2)}) = (\hat{a}^*_{\sigma,-}(f^{(1)})B^*(h'^{(1)}_\sigma)\Omega_0,$$

$$\hat{a}^*_{\sigma,-}(f^{(2)})B^*(h'^{(2)}_\sigma)\Omega_0)$$

$$+ i(\hat{a}^*_{\sigma,-}(f^{(1)})B^*(h'^{(1)}_\sigma)\Omega_0, \int_{-\infty}^{+\infty} dt_2 e^{it_2\hat{H}_\sigma} \mathfrak{A}^*_\sigma(f^{(2)};t_2)$$

$$\times e^{-it_2\hat{H}_\sigma} B^*(h'^{(2)}_\sigma)\Omega_0). \qquad (3.38)$$

Using again Theorem 3.5, the fact that the integral exists in the strong sense and *a fortiori* in the weak one, and $e^{-it_2\hat{H}_\sigma} B^*(h'^{(1)}_\sigma)\Omega_0 = B^*(e^{-it_2\omega} h'^{(1)}_\sigma)\Omega_0$ we have ("reduction from the right"):

$$\hat{S}_\sigma(h^{(1)}, f^{(1)}; h^{(2)}, f^{(2)}) = (h^{(1)}, h^{(2)})(f^{(1)}, f^{(2)})$$

$$+ i\int_{-\infty}^{+\infty} dt_2 (\hat{a}^*_{\sigma,-}(f^{(1)})B^*(h'^{(1)}_\sigma)$$

$$\times \Omega_0, e^{it_2\hat{H}_\sigma} \mathfrak{A}^*_\sigma(f^{(2)};t_2) B^*(e^{-it_2\omega} h'^{(2)}_\sigma)\Omega_0). \qquad (3.39)$$

Similarly, "reducing from the left,"

$$\hat{S}_\sigma(h^{(1)}, f^{(1)}; h^{(2)}, f^{(2)})$$

$$= (h^{(1)}, h^{(2)})(f^{(1)}, f^{(2)}) + \hat{S}^{(I)}_\sigma + \hat{S}^{(II)}_\sigma, \text{ with}$$

$$\hat{S}^{(I)}_\sigma \equiv i\int_{-\infty}^{+\infty} dt_2 F^{(1)}(t_2), F^{(1)}(t) \equiv (a^*(f^{(1)})B^*(h'^{(1)}_\sigma)\Omega_0,$$

$$e^{it\hat{H}_\sigma} \mathfrak{A}^*_\sigma(f^{(2)};t) B^*(e^{-it\omega} h'^{(2)}_\sigma)\Omega_0),$$

$$\hat{S}^{(II)}_\sigma \equiv \int_{-\infty}^{+\infty} dt_2 \Big(\int_0^{-\infty} dt_1 e^{it_1\hat{H}_\sigma} \mathfrak{A}^*_\sigma(f^{(1)};t_1)$$

$$\times B^*(e^{-it_1\omega} h'^{(1)}_\sigma)\Omega_0, e^{it_2\hat{H}_\sigma} \mathfrak{A}^*_\sigma(f^{(2)};t_2) B^*(e^{-it_2\omega} h'^{(2)}_\sigma)\Omega_0\Big).$$

We can now exploit the fact that the absolute value of the scalar product

$$(e^{it_1\hat{H}_\sigma} \mathfrak{A}^*_\sigma(f^{(1)};t_1) B^*(e^{-it_1\omega} h'^{(1)}_\sigma)\Omega_0, e^{it_2\hat{H}_\sigma} \mathfrak{A}^*_\sigma(f^{(2)};t_2)$$

$$\times B^*(e^{-it_2\omega} h'^{(2)}_\sigma)\Omega_0)$$

is, for fixed t_1, $O(|t_2|^{-3/2})$ as $|t_2| \to \infty$ and, for fixed t_2, $O(|t_1|^{-3/2})$ as $|t_1| \to \infty$, as shown using essentially Schwarz inequality, to conclude that

$$\hat{S}_\sigma^{(II)} = \int_{-\infty}^{+\infty} dt_2 \int_0^{-\infty} dt_1 F^{(II)}(t_2, t_1),$$

with

$$\begin{aligned} F^{(II)}(t_2,t_1) &\equiv (\mathfrak{A}_\sigma^*(f^{(1)};t_1)B^*(e^{-it_1\omega}h_\sigma'^{(1)})\Omega_0, e^{i(t_2-t_1)\hat{H}}{}_\sigma \mathfrak{A}_\sigma^*(f^{(2)};t_2) \\ &\quad \times B^*(e^{-it_2\omega}h_\sigma'^{(2)})\Omega_0) \\ &= (\mathfrak{A}_\sigma^*(f^{(1)};t_1)T_\sigma A_\sigma b^*(e^{-it_1\omega}h^{(1)})\Omega_0 \\ &\quad \times e^{i(t_2-t_1)\hat{H}}{}_\sigma \mathfrak{A}_\sigma^*(f^{(2)};t_2)T_\sigma A_\sigma b^*(e^{-it_2\omega}h^{(2)})\Omega_0). \end{aligned}$$

Let us remark that similar reduction formulas can also be obtained in the same way for the scattering of one physical nucleon and any finite number of mesons, using repeatedly the representation (3.37) for the asymptotic fields $\hat{a}_{\sigma,\pm}^*$.

Finally we would like to remark that the matrix elements $\hat{S}_\sigma(h^{(1)}, f^{(1)}; h^{(2)}, f^{(2)})$ (as well as the just mentioned ones) can also be expressed in terms of time ordered vacuum expectation values of the dressed fields (Green's functions). The sufficient smoothness conditions in t_1, t_2 of the functions can be established using, for $\sigma < \infty$, the higher order estimates we shall derive in part II, for $\sigma = \infty$ corresponding estimates of Ref. 16.

Note added to Proof: Since the writing of this paper new investigations of persistent models have appeared, concerned in particular with the solution of the one-body problem and the infrared problem. In particular: J. Fröhlich, "Existence of dressed one electron states in a class of persistent models," ETHZ (1972); L. Gross, J. Funct. Anal. **10**, 52 (1972); and "The relativistic polaron without cutoffs," Cornell (1972); D. Sloan, "The relativistic polaron without cutoffs in two space dimensions," Cornell (1971).

ACKNOWLEDGMENTS

It is a pleasure to thank Professor A. S. Wightman for the hospitality given to me at Princeton University. His enthusiastic attitude towards mathematical physics and his illuminating comments have had a very stimulating impact on my work.

I also thank Dr. R. Høegh-Krohn warmly for several very enlightening discussions about Nelson's model and its asymptotics. It is also a pleasure to thank Dr. J. P. Eckmann for clarifying discussions on his model and for a critical reading of the manuscript, as well as Professors M. Fierz, A. Jaffe, R. Jost, D. N. Williams, Drs. J. Chadam, H. Epstein, F. Guerra, J. Roberts, W. Schneider, R. Schrader, F. Strocchi for helpful discussions. I would like to thank also J. Fröhlich for kind information concerning his independent work on related questions and for interesting remarks.

Finally the financial support of the Swiss National Foundation and the Eidgenössische Technische Hochschule, Zürich is gratefully acknowledged. Part of this work was completed when I was a member of the Summer Institute in Theoretical Physics, 1971, at the University of Colorado (Boulder). I would like to thank Professors W. E. Brittin and A. O. Barut for their hospitality.

APPENDIX

In this Appendix we shall prove point 3 of Theorem 2.3, as well as the related propositions in Theorem 3.5. In the case $\sigma < \infty$ we can proceed as in Ref. 23. Take, e.g., the operators $\hat{a}_{\sigma,t}^{\#}(h)$ of Theorems 2.2, 2.3. We have, for any $\phi, \psi \in D(\hat{H}_0)$

$$\begin{aligned} (\psi, [a^{\#}(\bar{h}), a^{\#}(g)]\phi) &= (\psi, [\hat{a}_{\sigma,t}^{\#}(\bar{h}), \hat{a}_{\sigma,t}^{\#}(g)]\phi) \\ &= ((\hat{a}_{\sigma,t}^{\#}(\bar{h})^*\psi, \hat{a}_{\sigma,t}^{\#}(g)\phi) - ((\hat{a}_{\sigma,t}^{\#}(g)^*\psi, \hat{a}_{\sigma,t}^{\#}(\bar{h})\phi). \quad \text{(A1)} \end{aligned}$$

The right-hand side of (A1) converges to

$$((\hat{a}_{\sigma,\pm}^{\#}(\bar{h}))^*\psi, \hat{a}_{\sigma,\pm}^{\#}(g)\phi) - ((\hat{a}_{\sigma,\pm}^{\#}(g))^*\psi, \hat{a}_{\sigma,\pm}^{\#}(\bar{h})\phi).$$

If we can show that $\hat{a}_{\sigma,\pm}^{\#}(f)\phi \in D(\hat{a}_{\sigma,\pm}^{\#}(f_1))$ for any $f, f_1 \in \mathcal{D}$ then we are finished. This domain property is a consequence of a uniform bound in t of the form

$$\|\hat{a}_{\sigma,t}^{\#}(f)\hat{a}_{\sigma,t}^{\#}(f_1)\phi\| \leq C\|f\|\,\|f_1\|\,\|(\hat{H}_\sigma + b_\sigma)\phi\|, \quad \text{(A2)}$$

proved using (2.7), (2.16). Using (A2) one sees furthermore that $(\hat{a}_{\sigma,\pm}^{\#}(f)^*\psi, \hat{a}_{\sigma,\pm}^{\#}(f_1)\phi)$ is a bounded linear functional acting on ψ, hence $\hat{a}_{\sigma,\pm}^{\#}(f_1)\phi \in D(\hat{a}_{\sigma,\pm}^{\#}(f)^*)$, which ends the proof of the strong commutation relations for the $\hat{a}_{\sigma,\pm}^{\#}(h)$.

In the limit $\sigma \to \infty$ a simple uniform estimate of the form (A2) does not hold anymore, and above argument gives only the commutation relations in the weak form 3b of Theorem 2.3. This is also the form in which the commutation relations for the asymptotic meson fields in Nelson's model are given originally in Ref. 17. However, Höegh-Krohn himself has since improved his result to the proof of the commutation relations in the strong sense.[46] His proof applies also to our case and goes essentially as follows:

Set $\phi_t(\bar{h}) \equiv 2^{-1/2}(\hat{a}_t^*(h) + \hat{a}_t(\bar{h}))$, $\pi_t(g) = -i2^{-1/2}(\hat{a}_t^*(g) - \hat{a}_t(g))$, for $f, g \in \mathcal{D}$. These operators have $D(\hat{a}^{\#}(\bar{h}))$ resp. $D(a^{\#}(g))$ in their definition domains and are self-adjoint. The strong limits $\phi_\pm(\bar{h})$ resp. $\pi_\pm(g)$ for $t \to \infty$ exist because of Theorems 2.2, 2.3. $\phi_\pm, \pi_\pm$ are self-adjoint because $\phi_\pm(\bar{h}) - \phi_{t=0}(\bar{h})$ and $\pi_\pm(g) - \pi_{t=0}(g)$ are bounded and $\phi_{t=0}(\bar{h}), \pi_{t=0}(g)$ are self-adjoint. From what we have shown above, $\phi_\pm(\bar{h}), \pi_\pm(g)$ satisfy the canonical commutation relations in the weak form, on vectors ϕ, ψ as in Theorem 2.3, 3b. From this it follows by known methods[47] that $\phi_\pm(\bar{h}), \pi_\pm(g)$ satisfy the strong commutation relations (2.13) on a dense set.

*Supported by the Swiss National Foundation (Forschungsstipendium of the Eidgenössische Technische Hochschule, Zürich).

†Present address: Institute of Mathematics, University of Oslo, Blindern, Oslo 3, Norway.

[1] S. Albeverio, Ann. Phys. (N.Y.) **71**, 167 (1972).

[2] See, e.g., the discussions in H. Eckstein, Nuovo Cimento **4**, 1017 (1956). O. W. Greenberg, Ph.D. thesis (Princeton University, 1956); R. Haag, Phys. Rev. **112**, 669 (1958).

[3] See R. Haag, Ref. 2.

[4] See, e.g., R. Streater and A. S. Wightman, *PCT, Spin and Statistics, and All That* (Benjamin, New York, 1964); R. Jost, *The General Theory of Quantized Fields* (American Mathematical Society, Providence, R.I., 1965).

[5] Haag–Ruelle scattering theory: see R. Haag, Ref. 2, D. Ruelle, Helv. Phys. Acta **35**, 147 (1962). See also R. Jost, Ref. 4, and K. Hepp, in *Axiomatic Field Theory, 1965, Brandeis University Summer Institute in Theoretical Physics* (Gordon and Breach, New York, 1966), Vol. 1, p. 135.

[6] A basic postulate of LSZ theory. See Ref. 7.

[7] See, e.g., O. Steinmann, *Perturbations Expansions in Axiomatic Field Theory, Lecture Notes in Physics, 11* (Springer-Verlag, Berlin, 1970).

[8] See, e.g., K. Hepp, Ref. 5. For proofs in the related Haag-Araki framwork see Ref. 9.

[9] H. Araki and R. Haag, Commun. Math. Phys. **4**, 77 (1967).

[10]See, e.g., J. Glimm and A. Jaffe, Lectures given at the Summer School in les Houches (1970).
[11]References for other types of models can be found, e.g., in S. Albeverio, "Strong asymptotic convergence of states and fields in models of quantum field theory," Princeton University, June, Oct., 1971.
[12]See, e.g., Ref. 10.
[13]See below for references.
[14]E. Nelson, J. Math. Phys. **5,** 1190 (1964).
[15]J. T. Cannon, J. Funct. Anal. **8,** 101 (1971).
[16]A recent interesting investigation of Nelson's model is due to J. Fröhlich, E.T.H. thesis. Results are given and discussed in: "Mathematical discussion of models with persistent vacuum; the infrared problem," E.T.H. preprint, July 1971.
[17]R. Høegh-Krohn, J. Math. Phys. **11,** 185 (1970).
[18]Other results on Nelson's model are implied by our treatment of Eckmann's model.
[19]The interaction is clearly nonlocal because only the part $\Phi_b^+(x)\Phi_b^-(x)$ of the local term: Φ_b^2: is taken into consideration. The interaction is, however, translation invariant, which makes the scattering interesting with other kind of difficulties than in other more singular but nontranslation invariant models. Despite the nonlocality, one has some localization properties, as shown, e.g., from the existence of an S matrix with cluster properties (see below and J. Fröhlich, Ref. 16).
[20](a) J. P. Eckmann, Commun. Math. Phys. **18,** 247 (1970). (b)See also J. P. Eckmann, thesis (University of Geneva, 1970).
[21]K. Hepp, *Théorie de la renormalization*, Lecture Notes in Physics, Vol. 2 (Springer-Verlag, Berlin, 1969).
[22]"Adjusted" refers to Friedrichs' terminology. See e.g. K. O. Friedrichs, *Perturbation of Spectra in Hilbert Space*, Lectures in Appl. Math., Vol. III (Am. Math. Soc., Providence, R.I., 1965).
[23]See, e.g., (a) Y. Kato and N. Mugibayaski, Prog. Theor. Phys. **30,** 103 (1963); Prog. Theor. Phys. **45,** 628 (1971). (b)R. Høegh-Krohn, Commun. Math. Phys. **18,** 109 (1970); Commun. Math. Phys. **21,** 256 (1971). And references given therein.
[24]The terminology is explained at the end of Sec. 1.
[25]We recover by this a result of Eckmann (Ref. 20) and provide also a substitute for Theorem 13 in Ref. 20(a) (the statement of which is, at it stands, not completely precise).
[26]S. Albeverio, Helv. Phys. Acta **45,** 303 (1972) (Fierz Festschrift).
[27]An alternative procedure, using a direct study of vacuum expectation values, has been pursued by J. Fröhlich (Ref. 16).
[28]L. Rosen, Commun. Pure Appl. Math. **24,** 417 (1971).
[29]E.g., R. Jost, Ref. 4. R. Jost, Helv. Phys. Acta **39,** 21 (1966). S. Nelson, Proc. Am. Math. Soc. **27,** 110 (1971).
[30]A similar observation is, e.g., known from the study of Lee-type models.
[31]The adaptions are necessary when discussing asymptotic quantities, essentially because of the weaker decay in time of smooth solutions of the Klein–Gordon equation in lower space dimensions. The best possible decay rate for a general class of solutions is $\sup_x |\varphi(x,t)| \leq C|t|^{-s/2}$, with some constant C (see S. Nelson, Ref. 29).
[32]This implies the independence of most results on the statistics of the nucleons.
[33]Anticipating the more precise definitions, we shall give later, we call "dressed" fields the interacting fields having the property of creating physical one-particle states when applied to the vacuum. Heisenberg picture "adjusted" fields (see below) give for asymptotic times the so-called asymptotic physics fields (which act, by definition, as free fields on asymptotic physical states). Clearly asymptotic (physical) field act in the same way as dressed fields on the vacuum, but they act, in general, differently on general states. Similarly, products of interacting dressed fields create from the vacuum "dressed states," products of asymptotic fields "physical states." Dressed one-particle states are the same as physical one-particle states.
[34]T. Kato, *Perturbation Theory for Linear Operators* (Springer, New York, 1966).
[35]The physical one nucleon energy $E_\lambda(q)$ depends, for $\lambda \neq 0$, on q in a way different from $(m_b^2 + q^2)^{1/2}$.
[36]E.g., the choice of the "simplified model" or the choice which shall be discussed further from Sec. 3 on.
[37]For $n = 0$ one has of course the usual commutation relations for the $a^\#(h)$. The notation $\Delta^{\circ(n)}_\sigma$ is modelled on the one we shall find convenient to use later, when discussing also nucleon fields.
[38]This is so already in Lee models in lower space dimensions and with ultraviolet cut-off: See K. Hepp, Ref. 21, Chap. III. Note, however, that the situation in the Lee model is much simpler than in Eckmann's model, since in the former a second-order mass renormalization term is sufficient to yield the correct one-particle spectrum.
[39]For more details in relations with Wick monomials and related objects (including "Friedrichs graphs or diagrams") see, e.g., K. O. Friedrichs, Ref. 22, Chap. III, and K. Hepp, Ref. 21, Chap. I.
[40]The normalization function $\nu_\sigma(q)$ corresponds to the field strength renormalization constant Z of local relativistic quantum field theory. This is a situation well known from the Lee model with relativistic kinematics [see, e.g., F. Guerra, Nuovo Cimento A **68,** 258 (1970)].
[41]O. E. Lanford III, Ph.D. thesis (Princeton University, 1966).
[42]We shall use the same symbols for operators in $\mathscr{H}^{(n)}$ and $\mathscr{H}$, whenever we feel that no confusion should arise.
[43]The above theorem shows that the wave operators are partial isometries. As is well known from nonrelativistic scattering theory (see, e.g., the references given in Ref 1) equality of the ranges $\theta^{(n)}_{(\sigma),\pm}$ of the wave operators in the sector $\mathscr{H}^{(n)}$ is equivalent, in the case in which there is only the channel with free physical nucleons and mesons, with the unitarity of the theory in the sector $\mathscr{H}^{(n)}$ (unitarity of the S matrix defined from the wave operators: see Sec. 3B below). Asymptotic completeness is, even in the case in which the only channel is the one with free physical nucleons and mesons, a stronger condition, and requires, in this case (since $\hat{H}_\sigma$ has purely continuous spectrum in all $\mathscr{H}^{(n)}$, $n > 0$, due to translation invariance), equality of above ranges with the entire $\mathscr{H}^{(n)}$. In $\mathscr{H}^{(0)}$ the unitary and asymptotic completeness are, of course, trivial. For all other sectors these properties are certainly nontrivial [cf. the situation in nonrelativistic N-body theories and in Lee models: e.g., Ref. 1, R. Schrader, Commun. Math. Phys. **10,** 155 (1968)].
[44]The fact that the operators $\Omega_\sigma(t)$ are unbounded in $\mathscr{H}^{(n)}$ gives the motivation for our remark in Footnote 25. For an alternative approach to the convergence to scattering states see Ref. 16.
[45]In contrast to the operator $\hat{S}_\sigma$ for $\sigma < \infty$ which could, in principle, be unitary and even asymptotic complete (see footnote 44), we do not expect $\hat{S}_\infty$ to be unitary in $\mathscr{H}^{(n)}(R)$, $n > 0$, since nucleons coming in with momenta bounded by R can be scattered to momenta bigger than R.
[46]R. Høegh-Kohn, private communication, which is gratefully acknowledged.
[47]See, e.g., H. G. Tillmann, Acta Sci. Math. **24,** 258 (1963).

A restricted Bäcklund transformation*

David W. McLaughlin

Department of Mathematics, Iowa State University, Ames, Iowa 50010

Alwyn C. Scott

Department of Electrical and Computer Engineering, The University of Wisconsin, Madison, Wisconsin 53706

(Received 5 February 1973)

The Bäcklund transformation provides a mathematical tool which displays the interaction of solitons. Here a simple, systematic Bäcklund formalism is introduced which permits the explicit construction of these transformations for a restricted class of nonlinear wave equations. Traditionally a Bäcklund transformation has been viewed as a transformation of a solution surface of a partial differential equation into another surface which may not satisfy the same equation. In the present paper the term "restricted Bäcklund transformation" (hereafter abbreviated R-B) is used to refer to the case in which the transformed surface does satisfy the same equation. This formalism clarifies the nature of those transformations which have already been used to study nonlinear interactions in many physical problems. The formalism is introduced through a form of the linear Klein–Gordon equation. For this linear example a complete set of Fourier components is generated by a sequence of R-B transformations. This concrete example also indicates the type of results one can expect in the nonlinear case. For the nonlinear equation $\phi_{xy} = F(\phi)$, a theorem is established which states that R-B transformations exist if and only if the nonlinearity $F(\cdot)$ satisfies $F'' = \kappa F$, where κ is a constant. For such nonlinearities, the R-B transformations are explicitly constructed and are used to display exact nonlinear interactions. A relationship between the condition $F'' = \kappa F$, the existence of an infinite number of conservation laws, and the transformation theory is briefly discussed.

1. INTRODUCTION

One of the most startling phenomena in nonlinear physics is the existence of "solitons," remarkably stable pulselike solutions of certain nonlinear, dispersive wave equations. These solitons are so stable that they retain their identity after experiencing fully nonlinear interactions. Solitons have been observed in a wide range of physical phenomena (Ref. 34) in shallow water waves,[1] in hydromagnetic waves,[2] in ion-acoustic plasma waves,[3] in magnetic flux propagation in a Josephson junction,[4] in a nonlinear lattice.[5] Solitons form the basis for a description of lossless propagation of ultrashort laser pulses in an active medium.[6] Solitons have been observed experimentally[1,6,10] and numerically.[7] Computer movies have been made which clarify their interactions (available from Bell Telephone Laboratories).[8,9] Scott[10] has built a mechanical model which displays these solitons. They also arise in certain model nonlinear field theories.[3] Surveys of all these phenomena may be found in references.[6,10–12,34]

Not only have these solitons been observed experimentally, but explicit analytical expressions which represent them undergoing nonlinear interactions have been found. These expressions are exact solutions of the nonlinear, dispersive wave equations which model the physics. Such solutions have been found for the sine-Gordon equation,[6,13,14] the Korteweg de-Vries equation,[15–17] and a nonlinear Schrödinger equation.[18]

Lamb[6] discovered these solutions for the sine-Gordon equation,

$$\phi_{xy} = \sin\phi, \tag{1.1}$$

by applying the transformation theory of Bäcklund. Following the work of Seeger, Donth, and Kochendörfer,[14] he uses this transformation theory to construct an analytical expression for an N-soliton solution. In addition, Whitham[19] has shown that the Cole-Hopf transformation is of Bäcklund type, and Loewner[20,21] Rogers,[22–24] and Power, Rogers, and Osborne,[25] have used these transformations in the area of gas dynamics.

In view of these successes of the transformation theory, one immediately thinks of many questions. Can the Bäcklund theory be used as a tool to study other nonlinear physical phenomena? Given a nonlinear wave equation, how can one construct the transformation? Can the transformation's relationship to solitons be clarified? Several nonlinear equations which are known to possess solitons are also known to possess an infinite number of conservation laws, two of which are momentum and energy. Is there a relationship between the Bäcklund theory and these higher conservation laws? Can the theory provide a geometrical interpretation of these conservation laws?

We have obtained at least partial answers to all of these questions. This paper covers the first few. A second paper, now in preparation, discusses those questions concerning conservation laws.

In attacking these questions, we found it necessary to develop a clear understanding of the Bäcklund theory. The transformations of Bäcklund were designed to transform solution surfaces of second order partial differential equations with two independent variables into other surfaces. A general theory for such transformations was introduced by Bäcklund[26,27] as early as 1876, and subsequently studied by Clairin,[28] Goursat,[29] and Forsyth.[30] Eisenhart[31] discusses these transformations in his book on classical differential geometry. Unfortunately, this literature is difficult to understand. No doubt, this is partially because the language is currently "out of style." However, we believe a more fundamental source of the difficulty is the authors' insistence to work with such general second order equations. This generality prohibits them from obtaining concrete results.

Here we restrict ourselves to equations which may be put in the form

$$\phi_{xy} + \alpha\phi_x + \beta\phi_y = F(\phi), \qquad \alpha, \beta \text{ const.} \tag{1.2}$$

where the function $F(\cdot)$ does not depend explicitly on x or y. This restriction permits a rather simple Bäcklund

formalism. The purpose of this paper is twofold: first, to introduce, for a restricted class of Bäcklund transformations, a view point which permits a particularly direct and systematic construction of these transformations; and second, to use this approach in the study of equations of type (1.2).

Stated somewhat informally, our main results establish that a transformation of R-B type (see Sec. 2 for the definition of "R-B type") exists for equations of the form (1.2) if and only if

$$\text{(a)}\quad F(\cdot) \quad \text{is linear, or}$$

$$\text{(b)}\quad \alpha = \beta = 0 \quad \text{and} \quad \frac{d^2F}{d\phi^2} = \kappa F, \tag{1.3}$$

κ an arbitrary constant.

For equations (1.2) satisfying (1.3), we explicitly construct the R-B transformations. These transformations take the form of a pair of coupled, first order partial differential equations. We then establish the fundamental step in the application of these transformations to the study of interacting waves: namely, that certain triples of solutions of this pair of equations generate fourth solutions. An explicit formula prescribing this fourth solution in terms of the other three is presented.

It is interesting to notice that the same condition (1.3) has arisen in Kruskal's studies[32] as a sufficient condition for the existence of an infinite number of polynomial conservation laws. While Kruskal has not proven (1.3) necessary, he has been unable to find an infinite number of conservation laws for other functions $F(\cdot)$. Another intriguing feature is that, for linear $F(\cdot)$, the R-B transformations generate all possible Fourier components, a complete set of solutions. In Sec. 4F we present some comments concerning relationships between R-B transformations, conservation laws, completeness, and the "inverse method" of Gardner, Greene, Kruskal, and Muira[15] and Lax.[16] These comments form the basis of a second paper now in preparation. In this paper, we discuss the Korteweg-de Vries equation through Bäcklund techniques. Since the KdV equation is third order, an extension of the transformation theory is necessary.

Underlying our entire work is a view of the Bäcklund transformation which differs from the traditional one. This viewpoint leads us to define the R-B transformation itself as the solution of a coupled pair of partial differential equations. This pair may be studied with analytical techniques, rather than by traditional geometric methods.

In Sec. 2 we introduce the Bäcklund transformation, define the R-B class for Eqs. (1.2), and obtain the coupled pair of equations which the R-B transformation satisfies. In Sec. 3 we investigate the example $F(\cdot)$ linear. This example serves as a convenient vehicle through which we introduce several concepts which are important for more general $F(\cdot)$. Because it is so explicit, this linear example is the key to understanding our Bäcklund theory. The main results concerning (1.2) are presented in Sec. 4 and 5.

2. DEFINITIONS OF THE R-B Transformations

Traditionally the Bäcklund transformation for (1.2) is viewed as a transformation of a solution surface $\phi_{n-1}(x, y)$ of (1.2) into another surface $\phi_n(x, y)$, a surface which may or may not solve (1.2). Here we *demand* that both ϕ_{n-1} and ϕ_n solve (1.2). Thus, we view the Bäcklund transformation as a transformation between the solution surfaces, ϕ_{n-1} and ϕ_n, of the form

$$\phi_{n,x} = P(\phi_n; \phi_{n-1}; \phi_{n-1,x}; \phi_{n-1,y}; x; y), \tag{2.1a}$$

$$\phi_{n,y} = Q(\phi_n; \phi_{n-1}; \phi_{n-1,x}; \phi_{n-1,y}; x; y). \tag{2.1b}$$

[The subscripts $n-1$ and n indicate that we shall be considering several pairs of solutions of (2.1), indexed by n. Throughout this paper the functions (P, Q) are assumed to be sufficiently smooth to permit all required differentiation.]

Notice that if the transformations (P, Q) can be found, we have replaced the second order equation (1.2) for ϕ_n with a pair of coupled first order partial differential equations for ϕ_n. Moreover, this pair of equations depends explicitly on an auxiliary solution ϕ_{n-1}. We shall interpret ϕ_{n-1} as a known or "old" solution from which the "new" solution ϕ_n is generated by a Bäcklund transformation. We remark that for fixed ϕ_{n-1} not all solutions of (1.2) will solve (2.1). In fact, one of the purposes of this work is to understand more clearly the relationships of ϕ_n to ϕ_{n-1}.

Since Eq. (1.2) does not depend explicitly on x or y, one would prefer that the transformation (P, Q) not depend explicitly on x and y either. With this restriction and the notation

$$\tau \equiv \phi_n, \quad \xi_1 \equiv \phi_{n-1}, \quad \xi_2 \equiv \phi_{n-1,x}, \quad \xi_3 \equiv \phi_{n-1,y},$$
$$\xi_4 \equiv \phi_{n-1,xy}, \quad \xi_5 \equiv \phi_{n-1,xx}, \quad \xi_6 \equiv \phi_{n-1,yy}, \tag{2.2}$$

Eqs. (2.1) take the form

$$\phi_{n,x} \equiv P(\tau, \xi_1, \xi_2, \xi_3), \tag{2.3a}$$

$$\phi_{n,y} \equiv Q(\tau, \xi_1, \xi_2, \xi_3). \tag{2.3b}$$

To define (P, Q) we first insure that ϕ_n satisfies (1.2). To this end we differentiate (2.3a, b), use (2.2), and obtain two expressions for $\phi_{n,xy}$,

$$\phi_{n,xy} = Q\frac{\partial P}{\partial \tau} + \xi_3\frac{\partial P}{\partial \xi_1} + \xi_4\frac{\partial P}{\partial \xi_2} + \xi_6\frac{\partial P}{\partial \xi_3}, \tag{2.4a}$$

$$\phi_{n,xy} = P\frac{\partial Q}{\partial \tau} + \xi_2\frac{\partial Q}{\partial \xi_1} + \xi_5\frac{\partial Q}{\partial \xi_2} + \xi_4\frac{\partial Q}{\partial \xi_3}. \tag{2.4b}$$

Equations (2.4) together with (1.2) yield a pair of equations for (P, Q),

$$Q\frac{\partial P}{\partial \tau} + \xi_3\frac{\partial P}{\partial \xi_1} + \xi_4\frac{\partial P}{\partial \xi_2} + \xi_6\frac{\partial P}{\partial \xi_3} + \alpha P + \beta Q = F(\tau), \tag{2.5a}$$

$$P\frac{\partial Q}{\partial \tau} + \xi_2\frac{\partial Q}{\partial \xi_1} + \xi_5\frac{\partial Q}{\partial \xi_2} + \xi_4\frac{\partial Q}{\partial \xi_3} + \alpha P + \beta Q = F(\tau). \tag{2.5b}$$

Before inquiring into the meaning of (2.5), let us impose the condition that ϕ_{n-1} satisfy (1.2). This condition forces a relation between the variables $(\xi_1, \xi_2, \xi_3, \xi_4)$,

$$\xi_4 + \alpha\xi_2 + \beta\xi_3 = F(\xi_1). \tag{2.6}$$

Using (2.6) to eliminate ξ_4 from (2.5), we obtain

$$Q\frac{\partial P}{\partial \tau} + \xi_3\frac{\partial P}{\partial \xi_1} + [F(\xi_1) - \alpha\xi_2 - \beta\xi_3]\frac{\partial P}{\partial \xi_2} + \xi_6\frac{\partial P}{\partial \xi_3} + \alpha P + \beta Q = F(\tau). \tag{2.7a}$$

$$P\frac{\partial Q}{\partial \tau} + \xi_2\frac{\partial Q}{\partial \xi_1} + \xi_5\frac{\partial Q}{\partial \xi_2} + [F(\xi_1) - \alpha\xi_2 - \beta\xi_3]\frac{\partial Q}{\partial \xi_3} + \alpha P + \beta Q = F(\tau). \tag{2.7b}$$

Notice that $\xi_6 \frac{\partial P}{\partial \xi_3}$ and $\xi_5 \frac{\partial Q}{\partial \xi_2}$ are troublesome terms in (2.7). Clearly if these two terms were not present, (2.7) could be treated as a pair of coupled partial differential equations in four independent variables $(\tau, \xi_1, \xi_2, \xi_3)$ for the transformations $P = P(\tau, \xi_1, \xi_2, \xi_3)$ and $Q = Q(\tau, \xi_1, \xi_2, \xi_3)$. For this reason we restrict the class of transformations (P, Q) so that P is independent of ξ_3 and Q independent of ξ_2. Under this restriction the differential equations (2.7) reduce to

$$Q \frac{\partial P}{\partial \tau} + \xi_3 \frac{\partial P}{\partial \xi_1} + [F(\xi_1) - \alpha \xi_2 - \beta \xi_3] \frac{\partial P}{\partial \xi_2} + \alpha P + \beta Q = F(\tau), \quad (2.8a)$$

$$P \frac{\partial Q}{\partial \tau} + \xi_2 \frac{\partial Q}{\partial \xi_1} + [F(\xi_1) - \alpha \xi_2 - \beta \xi_3] \frac{\partial Q}{\partial \xi_3} + \alpha P + \beta Q = F(\tau), \quad (2.8b)$$

$$P = P(\tau, \xi_1, \xi_2), \qquad Q = Q(\tau, \xi_1, \xi_3).$$

Thus we are led to the following definition: Any (sufficiently differentiable) solution (P, Q) of (2.8) will be referred to as a *restricted Bäcklund (R–B) transformation mapping solutions of (1.2) into solutions of (1.2)*. Any such R-B transformation, evaluated at $\tau = \phi_n(x, y)$, $\xi_1 = \phi_{n-1}(x, y)$, $\xi_2 = \phi_{n-1,x}$, $\xi_3 = \phi_{n-1,y}$ and substituted into (2.3) results in the pair of first order equations

$$\phi_{n,x} = P(\phi_n; \phi_{n-1}; \phi_{n-1,x}), \quad (2.9a)$$

$$\phi_{n,y} = Q(\phi_n; \phi_{n-1}; \phi_{n-1,y}), \quad (2.9b)$$

where ϕ_{n-1} is any solution surface of (1.2). By the definition of (P, Q), any solution ϕ_n of this pair also satisfies (1.2).

Before continuing we remark that a significant simplification occurs if one further restricts (P, Q) to be independent of τ (i.e., ϕ_n). Firstly, Eqs. (2.8) which define the transformations are linear under this restriction. Secondly, if (P, Q) exist under this additional restriction, then Eqs. (2.9) express $\phi_{n,x}$ and $\phi_{n,y}$ completely in terms of known functions ϕ_{n-1}, $\phi_{n-1,x}$, $\phi_{n-1,y}$. Viewed in this manner, Forsyth's choice to restrict his study primarily to this case takes advantage of these simplifications. Unfortunately, as will become clear in Secs. 3 and 4, this is not sufficiently general for our purposes.

Finally, it should be noted that this class of R-B transformations is quite restrictive. This is particularly apparent from Eqs. (2.9). Nevertheless, as far as we know, only transformations of this form have been successfully utilized in the study of nonlinear waves. Therefore, it seems appropriate to investigate their degree of generality.

3. $F(\cdot)$ LINEAR

A. Transformations from the "vacuum"

In this section we study the case $F(\cdot)$ linear:

$$\phi_{xy} + \alpha \phi_x + \beta \phi_y = F(\phi) = \gamma \phi + \Gamma, \qquad \alpha, \beta, \gamma, \Gamma \text{ const.}, \quad \gamma \neq 0. \quad (3.1)$$

For an alternate view of linear problems, we refer the reader to the work of Loewner.[20,21]

As a first step in solving (2.8) for linear $F(\cdot)$, we seek solutions (P_0, Q_0) which are independent of (ξ_1, ξ_2, ξ_3). These functions must satisfy the coupled pair of ordinary differential equations

$$Q_0 \frac{dP_0}{d\tau} + \alpha P_0 + \beta Q_0 = \gamma \tau + \Gamma, \quad (3.2a)$$

$$P_0 \frac{dQ_0}{d\tau} + \alpha P_0 + \beta Q_0 = \gamma \tau + \Gamma, \quad (3.2b)$$

As long as $\gamma \neq 0$, (P_0, Q_0) generate a new solution from the critical or equilibrium solution $\phi_0 = -\Gamma/\gamma$ (or $\xi_1 = -\Gamma/\gamma$, $\xi_2 = \xi_3 = 0$). We remark that for the Klein-Gordon equation, this trivial solution is zero and is referred to as the "vacuum solution." Transformations from specific solutions such as the vacuum solution are not as general as transformations from an arbitrary solution ϕ_{n-1}, but they are more flexible.

To solve (3.2) for (P_0, Q_0), we subtract (3.2b) from (3.2a) and obtain

$$Q_0 \frac{dP_0}{d\tau} = P_0 \frac{dQ_0}{d\tau}, \quad (3.3)$$

which implies

$$Q_0 = c_0 P_0. \quad (3.4)$$

Here c_0 is a (possibly complex) constant which, at this point, is arbitrary. Together Eqs. (2.3) and (3.4) show that a solution generated from the "vacuum" will be a progressive (generalized traveling) wave of the form

$$\phi_1 = \phi_1(x + c_0 y). \quad (3.5)$$

Inserting (3.4) into (3.2) and solving for (P_0, Q_0), we find the particular solution

$$P_0 = l_0 + k_0 \tau, \quad (3.6a)$$

$$Q_0 = c_0(l_0 + k_0 \tau), \quad (3.6b)$$

where k_0 and c_0 are related by the *dispersion equation*

$$c_0 k_0^2 + (\alpha + \beta c_0) k_0 - \gamma = 0, \quad (3.7)$$

and l_0 is specified in terms of k_0 by the equation

$$l_0 = k_0 \Gamma/\gamma, \qquad \gamma \neq 0. \quad (3.8)$$

The dispersion equation (3.7) is naturally associated with the linear partial differential equation (3.1). Collecting these results we find

$$P_0 = k_0(\tau + \Gamma/\gamma), \quad (3.6a')$$

$$Q_0 = c_0 k_0(\tau + \Gamma/\gamma), \quad (3.6'b)$$

where c_0 and k_0 are related by (3.7).

Assuming $\gamma \neq 0$, we denote as ϕ_1 a solution generated from the vacuum by the R-B transformation (P_0, Q_0). Returning to (2.9), we write the first order partial differential equations for ϕ_1 as

$$\phi_{1,x} = k_0(\phi_1 + \Gamma/\gamma), \quad (3.9a)$$

$$\phi_{1,y} = c_0 k_0(\phi_1 + \Gamma/\gamma), \quad (3.9b)$$

from which we obtain

$$\phi_1(x, y) = -\Gamma/\gamma + A_0 e^{k_0(x + c_0 y)}. \quad (3.10)$$

Thus an R-B transformation from the vacuum generates an elementary progressive wave solution or Fourier component. The "amplitude" A_0 can be chosen arbitrarily, but the "wave number" k_0 and the "velocity" c_0 must be related through the dispersion equation (3.7).

Two features of the transformations (P_0, Q_0) should be emphasized. The first is the manner in which the dispersion equation arises in the transformation theory. The second is that a set of Fourier components may be generated which span any countable spectrum.

B. Transformation from an arbitrary solution

To generate ϕ_n from an arbitrary solution ϕ_{n-1}, it is sufficient for our purposes to seek transformations (P, Q) of the form

$$P(\tau, \xi_1, \xi_2) = k(\tau - \xi_1) + B\xi_2, \tag{3.11a}$$

$$Q(\tau, \xi_1, \xi_3) = ck(\tau - \xi_1) + D\xi_3. \tag{3.11b}$$

Here the constants c and k are again related by the dispersion relation (3.7), while the constants B and D are to be determined. This particular ansatz has been motivated by the transformations (P_0, Q_0), [(3.6), (3.7), (3, 8)], to which it reduces when ϕ_{n-1} is the "vacuum solution" ($\xi_1 = -\Gamma/\gamma$, $\xi_2 = 0$, $\xi_3 = 0$). Substituting (3.11) into (2.8), we obtain the equations

$$[ck^2 + k(\alpha + c\beta) - \gamma]\tau + [B\gamma - ck^2 - k(\alpha + c\beta)]\xi_1 + [\alpha\beta - \alpha\beta]\xi_2 + [k(D-1) + \beta(D - B)]\xi_3 = \Gamma(1 - B), \tag{3.12a}$$

$$[ck^2 + k(\alpha + c\beta) - \gamma]\tau + [D\gamma - ck^2 - k(\alpha + c\beta)]\xi_1 + [ck(B-1) + \alpha(B - D)]\xi_2 + [\beta D - \beta D]\xi_3 = \Gamma(1 - D). \tag{3.12b}$$

The coefficients of the independent variables must vanish individually. The coefficient of τ in both equations is just the dispersion relation (3.7). The remaining terms imply

$$B = D = 1. \tag{3.13}$$

Collecting these results, we obtain R-B transformations (P, Q),

$$P = k(\tau - \xi_1) + \xi_2, \tag{3.14}$$

$$Q = ck(\tau - \xi_1) + \xi_3, \tag{3.15}$$

where k and c satisfy the dispersion relation (3.7).

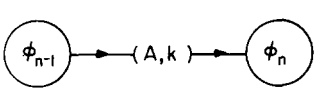


FIG. 1. Lamb diagram representing (3.16).

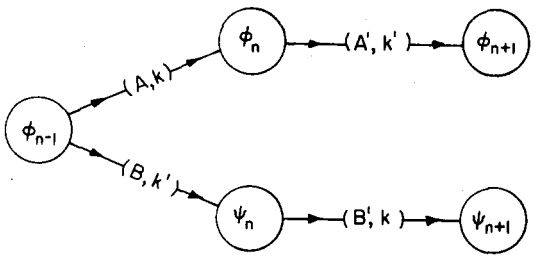


FIG. 2. Lamb diagram representing (3.21), (3.22), (3.23), and (3.24).

C. Relationship between R-B transformations and Fourier components

These R-B transformations (3.14), (3.15) when substituted into (2.9), yield a first order equation for ϕ_n in terms of any given solution surface ϕ_{n-1},

$$\phi_{n,x} = k(\phi_n - \phi_{n-1}) + \phi_{n-1,x}, \tag{3.16a}$$

$$\phi_{n,y} = ck(\phi_n - \phi_{n-1}) + \phi_{n-1,y}. \tag{3.16b}$$

Clearly a particular solution ϕ_n^p of this pair of equations is given by

$$\phi_n^p = \phi_{n-1}, \tag{3.17}$$

and the general solution by

$$\phi_n(x, y) = \phi_{n-1}(x, y) + Ae^{k(x+cy)}. \tag{3.18}$$

Here A is arbitrary, while c and k satisfy the dispersion equation (3.7).

Thus, the R-B transformation creates an arbitrary Fourier component, and adds it to ϕ_{n-1}. Assume for the moment, that boundary conditions have imposed a discrete spectrum. Further, think of ϕ_{n-1} as having been created by a succession of $(n-1)$ R-B transformations starting from the vacuum $\phi_0 = -\Gamma/\gamma$. Then ϕ_{n-1} is given by

$$\phi_{n-1}(x, y) = -\Gamma/\gamma + \sum_{j=1}^{n-1} A_j e^{k_j(x + c_j y)}, \tag{3.19}$$

where each k_j is a (distinct) member of the spectrum with the corresponding c_j defined by the dispersion relation. Clearly, when the spectrum is countable, one may get arbitrarily close to any solution by a sequence of R-B transformations beginning from the vacuum. Conversely, given any solution ϕ, it may be represented as a series of Fourier components,

$$\phi = \frac{-\Gamma}{\gamma} + \sum_{j=1}^{\infty} A_j e^{k_j(x + c_j y)} \tag{3.20}$$

and one may use a R-B transformation to annihiliate the lth component of ϕ by selecting $k = k_l$ and $A = -A_l$ in (3.16) and (3.18). By a sequence of such annihilations, one may get arbitrarily close to the vacuum solution from any given solution ϕ.

D. Lamb diagrams

Here we discuss, through this linear example, a schematic means to depict the R-B transformation. Consider the R-B transformation (3.14), (3.16), and (3.18) as a mapping of ϕ_{n-1} into ϕ_n. Clearly, it may be pictured by the diagram in Fig. 1, with the constants A and k explicitly shown. Diagrams such as Fig. 1 were introduced by Lamb[13] in his study of sine-Gordon equation; hence, the "Lamb diagram." These diagrams are convenient for recording and representing a series of R-B transformations, and they become particularly useful when studying the interaction of waves.

For this linear system wave interactions are trivial; but it is convenient to use this example to introduce an important method which employs Lamb diagrams to study interactions. Consider the Lamb diagram shown in Fig. 2. This diagram represents the following four pairs of equations and amplitudes (integration constants):

$$\phi_{n,x} = k(\phi_n - \phi_{n-1}) + \phi_{n-1,x}, \tag{3.21a}$$

$$\phi_{n,y} = ck(\phi_n - \phi_{n-1}) + \phi_{n-1,y}, \text{ amplitude } A; \tag{3.21b}$$

$$\phi_{n+1,x} = k'(\phi_{n+1} - \phi_n) + \phi_{n,x}, \tag{3.22a}$$

$$\phi_{n+1,y} = c'k'(\phi_{n+1} - \phi_n) + \phi_{n,y}, \quad \text{amplitude } A'; \tag{3.22b}$$

$$\psi_{n,x} = k'(\psi_n - \phi_{n-1}) + \phi_{n-1,x} \tag{3.23a}$$

$$\psi_{n,y} = c'k'(\psi_n - \phi_{n-1}) + \phi_{n-1,y}, \quad \text{amplitude } B; \tag{3.23b}$$

$$\psi_{n+1,x} = k(\psi_{n+1} - \psi_n) + \psi_{n,x}, \tag{3.24a}$$

$$\psi_{n+1,y} = ck(\psi_{n+1} - \psi_n) + \psi_{n,y}, \quad \text{amplitude } B'. \tag{3.24b}$$

Using (3.18) to solve for ϕ_{n+1} and ψ_{n+1} in terms of ϕ_{n-1}, we obtain

$$\phi_{n+1} = Ae^{k(x+cy)} + A'e^{k'(x+cy)} + \phi_{n-1}, \tag{3.25a}$$

$$\psi_{n+1} = Be^{k'(x+cy)} + B'e^{k(x+cy)} + \phi_{n-1}. \tag{3.25b}$$

Notice that if we choose $A' = B$ and $B' = A$, $\phi_{n+1} = \psi_{n+1}$. In retrospect we have observed that the Lamb diagram in Fig. 3 represents a valid R-B transformation.

Had we been given the fact, displayed in Fig. 3, that the Lamb diagram commutes, we could have found ϕ_{n+1} without explicitly integrating the partial differential equations (3.22). Equations (3.21a) and (3.22a) imply

$$\phi_{n+1,x} = k'(\phi_{n+1} - \phi_n) + k(\phi_n - \phi_{n-1}) + \phi_{n-1,x}, \tag{3.26}$$

while the Eqs. (3.23a) and (3.24a) imply

$$\psi_{n+1,x} = k(\psi_{n+1} - \psi_n) + k'(\psi_n - \phi_{n-1}) + \phi_{n-1,x}. \tag{3.27}$$

Setting $\psi_{n+1} = \phi_{n+1}$, as in Fig. 3, we obtain the algebraic relation among $\phi_{n-1}, \phi_n, \psi_n, \phi_{n+1}$,

$$(k - k')[\phi_{n+1} - \phi_n - \psi_n + \phi_{n-1}] = 0. \tag{3.28}$$

The "b equations" yield the same result.

Equation (3.28) gives no information if $k = k'$; however, if $k \neq k'$, we obtain the particular solution ϕ_{n+1},

$$\phi_{n+1} = \phi_n + \psi_n - \phi_{n-1}, \tag{3.29}$$

as we must since the interaction is linear.

For this linear case the observation that particular solutions ϕ_{n+1} exist which permit the Lamb diagrams to commute (Fig. 3) was made by an examination of the explicit solutions ϕ_{n+1} and ψ_{n+1}, (3.25a) and (3.25b). For (nonlinear) $F(\cdot)$ satisfying $F''(\cdot) = \kappa F(\cdot)$, we will give an indirect proof that the Lamb diagrams commute, and then use this fact to study interacting solutions.

We close this section with the reminder that when $\gamma = 0$ (i.e., $F = \Gamma$) the R-B transformation given by (3.14) still applies. However, the dispersion relation is now

$$ck^2 + (\alpha + \beta c)k = 0. \tag{3.30}$$

4. $\phi_{xy} = F(\phi)$

In this section we study the equation

$$\phi_{xy} = F(\phi), \tag{4.1}$$

where $F(\cdot)$ is an arbitrary smooth function. Since we have already treated the case $F(\cdot)$ linear, we exclude linear $F(\cdot)$ from this section. More precisely, we assume $d^2F(\phi)/d\phi^2$ does not vanish identically in any open interval under consideration. For $\phi_{xy} = F(\phi)$, Eqs. (2.8), which define the R-B transformations, become

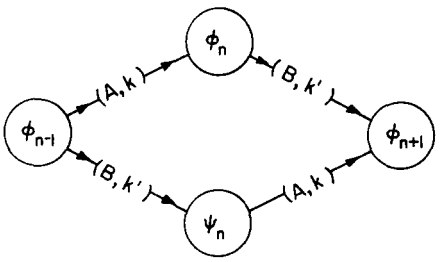


FIG. 3. Commuting Lamb diagram corresponding to Fig. 2.

$$Q\frac{\partial P}{\partial \tau} + \xi_3 \frac{\partial P}{\partial \xi_1} + F(\xi_1)\frac{\partial P}{\partial \xi_2} = F(\tau), \tag{4.2a}$$

$$P\frac{\partial Q}{\partial \tau} + \xi_2 \frac{\partial Q}{\partial \xi_1} + F(\xi_1)\frac{\partial Q}{\partial \xi_3} = F(\tau), \tag{4.2b}$$

$$P = P(\tau, \xi_1, \xi_2), \quad Q = Q(\tau, \xi_1, \xi_3).$$

A. Progressive waves from the "vacuum"

We begin our investigation of (4.2) by seeking solutions (P_0, Q_0) independent of (ξ_1, ξ_2, ξ_3). If $F(\cdot)$ has a zero ϕ_0, then (P_0, Q_0) may be considered as transformations which generate from the equilibrium or "vacuum" solution $\phi_0(\xi_1 = \phi_0, \xi_2 = \xi_3 = 0)$. Equations (4.2) show that (P_0, Q_0) satisfy

$$Q_0 \frac{dP_0}{d\tau} = F(\tau), \tag{4.3a}$$

$$P_0 \frac{dQ_0}{d\tau} = F(\tau). \tag{4.3b}$$

Equations (4.3) imply

$$Q_0 = c_0 P_0, \quad c_0 (\neq 0) \text{ const}, \tag{4.4}$$

which, from (2.9), required that $\phi_1 (\equiv \tau)$ be a progressive wave of the form

$$\phi_1 = \phi_1(x + c_0 y). \tag{4.5}$$

Substitution of (4.4) into (4.3) and integration yields

$$P_0(\tau) = 2a_0[\tfrac{1}{2}(E + G(\tau))]^{1/2} \tag{4.6a}$$

$$Q_0(\tau) = (2/a_0)[\tfrac{1}{2}(E + G(\tau))]^{1/2}, \quad a_0^2 = 1/c_0, \tag{4.6b}$$

where E is a constant of integration and

$$G(\tau) \equiv \int^{\tau} F(\tau')d\tau'. \tag{4.7}$$

Equations (4.6) and (2.9) imply that the progressive wave ϕ_1 must satisfy

$$\phi_{1,x} = 2a_0[\tfrac{1}{2}(E + G(\phi_1))]^{1/2}, \tag{4.8a}$$

$$\phi_{1,y} = (2/a_0)[\tfrac{1}{2}(E + G(\phi_1))]^{1/2}. \tag{4.8b}$$

Thus ϕ_1 is defined by the integral

$$\frac{1}{a_0}\int^{\phi_1}[2(E + G(\tau))]^{-1/2}d\tau = x + c_0 y + \theta, \tag{4.9}$$

θ constant of integration, $\quad c_0 = 1/a_0^2$.

Equation (4.9) yields all progressive waves of $\phi_{xy} = F(\phi)$ as may be easily verified by seeking solutions of

(4.1) of the form $\phi = \phi(x + c_0 y)$, $c_0 \neq 0$. If we transform to "space-time" variables (z, t) defined by $z + st = x$, $z - st = y$, $s > 0$ const, we see that the transformations (P_0, Q_0) generate all progressive waves traveling at constant speed $v = [(1 - c_0)/(1 + c_0)]s$. Thus any velocity v lying between the two characteristic velocities $\pm s$ may be obtained by adjusting $c_0 > 0$,

$$-s < \left(\frac{1 - c_0}{1 + c_0}\right)s \equiv v < s, \qquad c_0 \in (0, \infty). \tag{4.10}$$

When $F(\phi) = \sin\phi$, these progressive waves have been studied in some detail[4,10,13] Only the cases $E \geq 1$, $|v| < s$ are stable.[10] When $E = 1, \phi_1$ represents a single "kink" or soliton [a solution which satisfies $\lim_{\xi \to -\infty} \phi(\xi) = 0$ modulo 2π and which increases by 2π as $\xi \to +\infty$]. For $E > 1, \phi_0$ represents an infinite wave train of equally spaced solitons traveling at constant velocity. For $-1 < E < 1$, $|v| < 1$, ϕ_1 is periodic about π. For $-1 < E < 1$, $|v| > 1$, ϕ_1 is periodic about 0. It is important to realize that all of these progressive wave forms may be generated from the "vacuum" by the "vacuum" transformations (P_0, Q_0).

B. Existence of R-B transformation implies $F''(\phi) = \kappa F(\phi)$ (κ constant)

In this section we prove that $F(\cdot)$ must satisfy $F'' = \kappa F$ if R-B transformations are to exist. We state this result in the form of a theorem.

Theorem 1: Let $F(\tau)$ be a strictly nonlinear analytic function of τ. Define the class of R-B transformations to be those solutions of (4.2) which are analytic in each argument $(\tau, \xi_1, \xi_2, \xi_3)$. Further, restrict the transformations (P, Q) to depend upon at least one member of the set (ξ_1, ξ_2, ξ_3). If such (P, Q) exist, then $F(\cdot)$ must satisfy $F'' = \kappa F$, where κ is some arbitrary, nonvanishing constant.

This theorem shows that the class of R-B transformations associated with $\phi_{xy} = F(\cdot)$ is quite restrictive. In fact, one may transform between such equations with $F'' = \kappa F$ by transformations of the form $\phi \to d\phi + e$, d and e complex constants, together with appropriate limits as $e \to \pm\infty$. This restrictive class of F is consistent with a relationship of the R-B transformations to the existence of an infinite number of conservation laws. This relationship is discussed elsewhere.[34,35] We remark that the assumptions of analyticity are more restrictive than necessary as may be seen by examining the following proof. We made these assumptions only for convenience of "book-keeping." The remainder of this section is devoted to the proof of the theorem.

Throughout the proof of this theorem we use the symbol $\equiv$ to mean "identically equal to." We assume R-B transformations (P, Q) exist satisfying the conditions imposed in the hypothesis. Applying $\partial^2/\partial\xi_3^2$ to (4.2a) yields

$$\frac{\partial^2 Q}{\partial \xi_3^2} \cdot \frac{\partial P}{\partial \tau} \equiv 0. \tag{4.11}$$

P is not independent of τ, for if it were, Eq. (4.2a) would imply $F'(\tau) \equiv 0$. Therefore, $\partial^2 Q/\partial\xi_3^2 \equiv 0$, which upon integration yields

$$Q(\tau, \xi_1, \xi_3) = q_0(\tau, \xi_1) + q(\tau, \xi_1)\xi_3. \tag{4.12}$$

Similarly, the application of $\partial^2/\partial\xi_2^2$ to (4.2b) implies

$$P(\tau, \xi_1, \xi_2) = p_0(\tau, \xi_1) + p(\tau, \xi_1)\xi_2. \tag{4.13}$$

Thus, the ξ_2 dependence of P and the ξ_3 dependence of Q are at most linear. Substituting (4.12) and (4.13) into (4.2) yields

$$\left(q_0 \frac{\partial p_0}{\partial \tau} + F(\xi_1)p - F(\tau)\right) + \left(q_0 \frac{\partial p}{\partial \tau}\right)\xi_2 + \left(q \frac{\partial p_0}{\partial \tau} + \frac{\partial p_0}{\partial \xi_1}\right)\xi_3 + \left(q \frac{\partial p}{\partial \tau} + \frac{\partial p}{\partial \xi_1}\right)\xi_2\xi_3 \equiv 0, \tag{4.14a}$$

$$\left(p_0 \frac{\partial q_0}{\partial \tau} + F(\xi_1)q - F(\tau)\right) + \left(p \frac{\partial q_0}{\partial \tau} + \frac{\partial q_0}{\partial \xi_1}\right)\xi_2 + \left(p_0 \frac{\partial q}{\partial \tau}\right)\xi_3 + \left(p \frac{\partial q}{\partial \tau} + \frac{\partial q}{\partial \xi_1}\right)\xi_2\xi_3 \equiv 0. \tag{4.14b}$$

Notice that all ξ_2 and ξ_3 dependence is explicitly shown in (4.14). Equating coefficients of like powers of (ξ_2, ξ_3) yields the following eight equations which must be satisfied by the functions p_0, q_0, p, q:

$$q_0 \frac{\partial p}{\partial \tau} \equiv 0, \tag{4.15a}$$

$$p_0 \frac{\partial q}{\partial \tau} \equiv 0; \tag{4.15b}$$

$$q \frac{\partial p}{\partial \tau} + \frac{\partial p}{\partial \xi_1} \equiv 0, \tag{4.16a}$$

$$p \frac{\partial q}{\partial \tau} + \frac{\partial q}{\partial \xi_1} \equiv 0; \tag{4.16b}$$

$$q \frac{\partial p_0}{\partial \tau} + \frac{\partial p_0}{\partial \xi_1} \equiv 0, \tag{4.17a}$$

$$p \frac{\partial q_0}{\partial \tau} + \frac{\partial q_0}{\partial \xi_1} \equiv 0; \tag{4.17b}$$

and

$$q_0 \frac{\partial p_0}{\partial \tau} + F(\xi_1)p - F(\tau) \equiv 0, \tag{4.18a}$$

$$p_0 \frac{\partial q_0}{\partial \tau} + F(\xi_1)q - F(\tau) \equiv 0. \tag{4.18b}$$

Notice that q_0 cannot be identically zero, for, if $q_0 \equiv 0$, Eqs. (4.18) yield $p = F(\tau)/F(\xi_1)$ and $q = F(\tau)/F(\xi_1)$. But then (4.16a) implies $F'(\tau) \equiv F'(\xi_1)$, which is a contradiction since $F(\cdot)$ is not linear. Since q_0 is not identically zero, (4.15a) demands

$$\frac{\partial p}{\partial \tau} \equiv 0. \tag{4.19}$$

Equation (4.16a) then demands

$$\frac{\partial p}{\partial \xi_1} \equiv 0. \tag{4.20}$$

Equations (4.19) and (4.20) together imply p is a constant,

$$p \equiv \bar{p}, \quad \bar{p} \text{ const}, \tag{4.21a}$$

$$q_0 \not\equiv 0. \tag{4.21b}$$

Through an analogous argument, Eqs. (4.15b), (4.16b), and (4.18) yield

$$q \equiv \bar{q}, \quad \bar{q} \text{ const}, \tag{4.22a}$$

$$p_0 \not\equiv 0. \tag{4.22b}$$

In Appendix A we establish that $\bar{p} \neq \bar{q}$, $\bar{p} \neq 0$, $\bar{q} \neq 0$, $\partial p_0/\partial\tau \not\equiv 0, \partial q_0/\partial\tau \not\equiv 0, \partial p_0/\partial\xi_1 \not\equiv 0, \partial q_0/\partial\xi_1 \not\equiv 0$. In view of these facts, we define the (nonsingular) transformation of the independent variables

$$u \equiv \tau - \bar{p}\xi_1 \quad \text{or} \quad \tau = (\bar{q}u - \bar{p}v)/(\bar{q} - \bar{p}), \tag{4.23}$$
$$v \equiv \tau - \bar{q}\xi_1, \quad \xi_1 = (u - v)/(\bar{q} - \bar{p}).$$

In terms of variables u and v, (4.17a, b) and the lemma of Appendix A imply

$$p_0 = \tilde{p}_0(v), \tag{4.24a}$$

$$q_0 = \tilde{q}_0(u). \tag{4.24b}$$

Thus, the problem is reduced to the pair (4.18a, b), which may be written as

$$\tilde{q}_0\tilde{p}_0' \equiv F(\tau) - \bar{p}F(\xi_1), \tag{4.25a}$$

$$\tilde{p}_0\tilde{q}_0' \equiv F(\tau) - \bar{q}F(\xi_1), \tag{4.25b}$$

where τ and ξ_1 are given in (4.23).

Taking $\partial/\partial u$ of Eq. (4.25a) and equating the result to the $\partial/\partial v$ of Eq. (4.25b), we obtain the condition $(\bar{q} + \bar{p})(F'(\tau) - F'(\xi_1)) \equiv 0$. Since $F(\cdot)$ is not linear, $\bar{q} = -\bar{p}$. In terms of this constant ($\bar{q} \equiv c$), (4.25a, b) become

$$\tilde{q}_0\tilde{p}_0' \equiv F\left(\frac{u+v}{2}\right) + cF\left(\frac{u-v}{2c}\right), \tag{4.26a}$$

$$\tilde{p}_0\tilde{q}_0' \equiv F\left(\frac{u+v}{2}\right) - cF\left(\frac{u-v}{2c}\right), \tag{4.26b}$$

$$\tau = \frac{u+v}{2}, \quad \xi_1 = \frac{u-v}{2c}.$$

Clearly (4.26a, b) imply $\tilde{p}_0' \not\equiv 0$, $\tilde{q}_0 \not\equiv 0$. Thus, we are permitted to solve (4.26a) for $\tilde{q}_0$, calculate $\tilde{q}_0'$, and multiply by $\tilde{p}_0$ to obtain

$$\tilde{p}_0(v)\tilde{q}_0'(u) = \frac{\tilde{p}_0(v)}{2\tilde{p}_0'(v)}\left[F'\left(\frac{u+v}{2}\right) + F'\left(\frac{u-v}{2c}\right)\right]. \tag{4.27}$$

By Eq. (4.26b) we obtain the condition

$$\frac{\tilde{p}_0(v)}{2\tilde{p}_0'(v)}\left[F'\left(\frac{u+v}{2}\right) + F'\left(\frac{u-v}{2c}\right)\right] \equiv \left(F\ \frac{u+v}{2}\right) - cF\left(\frac{u-v}{2c}\right). \tag{4.28}$$

Taking $\partial/\partial u$ of (4.28) and eliminating $\tilde{p}_0/2\tilde{p}_0'$ yields

$$(F'(\tau))^2 - (F'(\xi_1))^2 + F(\xi_1)F''(\xi_1) - F(\tau)F''(\tau) + c[F(\xi_1)F''(\tau) - (1/c^2)F(\tau)F''(\xi_1)] \equiv 0. \tag{4.29}$$

Finally, taking $\partial^2/\partial\tau\partial\xi_1$ of (4.29), we find

$$\frac{F'''(\tau)}{F'(\tau)} \equiv \frac{1}{c^2}\frac{F'''(\xi_1)}{F'(\xi_1)}. \tag{4.30}$$

In Appendix B we establish that no solution (P, Q) exists for $F(\tau) = \Gamma + \gamma_1\tau + \gamma_2\tau^2, \gamma_2 \neq 0$. Therefore, $c^2 = 1$ and $F'''(\tau) = \kappa F'(\tau)$, $\kappa \neq 0$. Integrating this equation for $F''(\cdot)$, we see that for (P, Q) to exist,

$$F''(\tau) \equiv \kappa F(\tau) + \kappa_1, \quad \kappa_1 \text{ const}. \tag{4.31}$$

We demonstrate in Appendix C that $\kappa_1 = 0$. Thus the theorem is established.

C. *R-B* transformations for $F'' = \kappa F$, $\kappa \neq 0$

In this section we present the transformations (P, Q) where $F(\cdot)$ satisfies $F'' = \kappa F$, $\kappa \neq 0$. The formulas are displayed in Theorem 2.

Theorem 2: Let $F(\tau) = Ae^{\nu\tau} + Be^{-\nu\tau}$, $\nu^2 = \kappa \neq 0$, where the (complex) constants A and B satisfy $|A|^2 + |B|^2 > 0$. The most general *R-B* transformations satisfying the restrictions of Theorem 1 are given by

$$P(\tau, \xi_1, \xi_2) = \xi_2 + \frac{2a}{\nu}F[\tfrac{1}{2}(\tau + \xi_1)], \tag{4.32a}$$

$$Q(\tau, \xi_1, \xi_3) = -\xi_3 + \frac{2}{a}\sinh\left(\frac{\nu}{2}(\tau - \xi_1)\right), \tag{4.32b}$$

and

$$P(\tau, \xi_1, \xi_2) = -\xi_2 + \frac{2}{a}\sinh\left(\frac{\nu}{2}(\tau - \xi_1)\right), \tag{4.33a}$$

$$Q(\tau, \xi_1, \xi_3) = \xi_3 + \frac{2a}{\nu}F[\tfrac{1}{2}(\tau + \xi_1)], \tag{4.33b}$$

where a is an arbitrary constant.

To prove this theorem, it is sufficient to find solutions $(\tilde{p}_0, \tilde{q}_0)$ of (4.26) with $c = \pm 1$, for then the *R-B* transformations (P, Q) are given by

$$P(\tau, \xi_1, \xi_2) = -c\xi_2 + \tilde{p}_0(\tau - c\xi_1), \tag{4.34a}$$

$$Q(\tau, \xi_1, \xi_3) = c\xi_3 + \tilde{q}_0(\tau + c\xi_1), \quad c = \pm 1. \tag{4.34b}$$

Select $c = -1$. With $F(\tau) = Ae^{\nu\tau} + Be^{-\nu\tau}$, (4.26) takes the form

$$\tilde{q}_0(u)\tilde{p}_0'(v) = (e^{\nu u/2} - e^{-\nu u/2})(Ae^{\nu v/2} - Be^{-\nu v/2}), \tag{4.35a}$$

$$\tilde{p}_0(v)\tilde{q}_0'(u) = (e^{\nu u/2} + e^{-\nu u/2})(Ae^{\nu v/2} + Be^{-\nu v/2}), \tag{4.35b}$$

which integrate to yield

$$\tilde{q}_0(u)\tilde{p}_0(v) = \frac{2}{\nu}(e^{\nu u/2} - e^{-\nu u/2})(Ae^{\nu v/2} + Be^{-\nu v/2}) + \kappa_1, \quad \kappa_1 \text{ const}. \tag{4.36}$$

From (4.35a) and (4.36), $\tilde{p}_0'/\tilde{p}_0$ is independent of u only for $\kappa_1 = 0$. Then we can write

$$\tilde{p}_0(v) = \frac{2a}{\nu}F[\tfrac{1}{2}(\tau + \xi_1)], \tag{4.37a}$$

$$\tilde{q}_0(u) = \frac{2}{a}\sinh\left(\frac{\nu}{2}(\tau - \xi_1)\right), \quad a \neq 0, \text{an arbitrary constant}. \tag{4.37b}$$

Thus, (4.37) establishes (4.32). The proof of (4.33) proceeds analogously with $c = +1$. This completes the proof of Theorem 2.

Notice that, when viewed through the partial differential equations (2.9) for ϕ_n, (4.32) differs from (4.33) only by interchange of x and y. Therefore, we restrict our attention to (4.32).

Also notice that for $A \neq 0$, $B \neq 0$, the function $F(\cdot)$ has zero(s) τ_0 defined by

$$F(\tau_0) = 0 \Rightarrow e^{2\nu\tau_0} = \left|\frac{B}{A}\right| e^{i(\pi + \theta_B - \theta_A)}, \tag{4.38}$$

where θ_A and θ_B are the phases of A and B, respectively.

Thus $\phi = \tau_0$ is an equilibrium solution of $\phi_{xy} = F(\phi)$, and $[P(\tau, \tau_0, 0), Q(\tau, \tau_0, 0)]$ represent R-B transformations from this equilibrium solution. All such transformations are given by (P_0, Q_0), (4.16). In particular, selecting the parameter $E = \sqrt{-4AB}/\nu$ in (P_0, Q_0) to be unity shows that $[P(\tau, \tau_0, 0), Q(\tau, \tau_0, 0)]$ is a special case of (P_0, Q_0). This establishes the claim that the vacuum transformations (P_0, Q_0) are more versatile than the R-B transformations (P, Q) when transforming from the vacuum solution.

We now specialize to the case $F(\cdot) = \sin(\cdot)$. When transforming from the vacuum, $[P(\tau, 0, 0), Q(\tau, 0, 0)]$ generates a single "kink" traveling at "speed" $c = 1/a^2$. Thus, we see that R-B transformations (4.32) generate "single kink" or soliton solutions. When used to transform an arbitrary solution ϕ, these transformations "add" an additional soliton to the solution. In principle, R-B transformations may be used to build solutions from basic soliton components. This property of R-B transformations is most intriguing (see Sec. 4F).

D. Lamb diagrams

In Sec. 3C we took the R-B transformations (3.14) for the linear problem, inserted them into the differential equation (2.9) for ϕ_n, and integrated to find ϕ_n. In this nonlinear case we do not integrate explicitly the analogous differential equations for ϕ_n. Nevertheless, by using "Lamb diagrams," we obtain an explicit solution ϕ_{n+1} in terms of a certain triple of solutions.

From Eqs. (2.9) and (4.32), we obtain

$$\phi_{n,x} = \phi_{n-1,x} + \frac{2a}{\nu} F[\tfrac{1}{2}(\phi_n + \phi_{n-1})], \tag{4.39a}$$

$$\phi_{n,y} = -\phi_{n-1,y} + \frac{2}{a}\sinh\left(\frac{\nu}{2}(\phi_n - \phi_{n-1})\right), \qquad \nu^2 = \kappa \neq 0. \tag{4.39b}$$

Schematically, R-B transformations (4.39) may be represented by the Lamb diagram shown in Fig. 4 where (a, b) denotes the constant a of (4.39) and another constant of integration b which arises upon integration of (4.39).

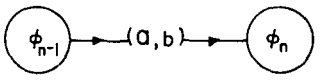


FIG. 4. Lamb diagram representing (4.39).

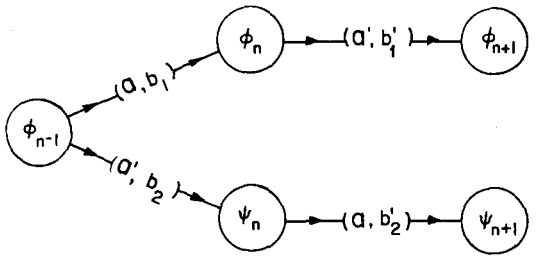


FIG. 5. Lamb diagram representing (4.40), (4.41), (4.42) and (4.43).

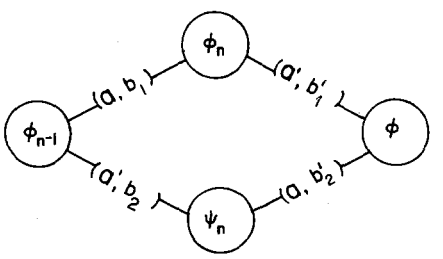


FIG. 6. Commuting Lamb diagram corresponding to Fig. 5.

Consider the Lamb diagram shown in Fig. 5. This diagram represents the following four pairs of equations and integration constants:

$$\phi_{n,x} = \phi_{n-1,x} + \frac{2a}{\nu} F[\tfrac{1}{2}(\phi_n + \phi_{n-1})], \tag{4.40a}$$

$$\phi_{n,y} = -\phi_{n-1,y} + \frac{2}{a}\sinh\left(\frac{\nu}{2}(\phi_n - \phi_{n-1})\right), \tag{4.40b}$$

integration constant b_1;

$$\phi_{n+1,x} = \phi_{n,x} + \frac{2a'}{\nu} F[\tfrac{1}{2}(\phi_{n+1} + \phi_n)], \tag{4.41a}$$

$$\phi_{n+1,y} = -\phi_{n,y} + \frac{2}{2'}\sinh\left(\frac{\nu}{2}(\phi_{n+1} - \phi_n)\right), \tag{4.41b}$$

integration constant b_1';

$$\psi_{n,x} = \phi_{n-1,x} + \frac{2a'}{\nu} F[\tfrac{1}{2}(\psi_n + \phi_{n-1})], \tag{4.42a}$$

$$\psi_{n,y} = -\phi_{n-1,y} + \frac{2}{a'}\sinh\left(\frac{\nu}{2}(\psi_n - \phi_{n-1})\right), \tag{4.42b}$$

integration constant b_2;

$$\psi_{n+1,x} = \psi_{n,x} + \frac{2a}{\nu} F[\tfrac{1}{2}(\psi_{n+1} + \psi_n)], \tag{4.43a}$$

$$\psi_{n+1,y} = \psi_{n,y} + \frac{2}{a}\sinh\left(\frac{\nu}{2}(\psi_{n+1} - \psi_n)\right), \tag{4.43b}$$

integration constant b_2'.

We now show that there exist solutions ψ_{n+1} and ϕ_{n+1} which are equal, i.e., such that Fig. 6 is valid. We have the following theorem.

Theorem 3: Consider an arbitrary solution ϕ_{n-1} of Eq. (4.1) with $F(\tau) = Ae^{\nu\tau} + Be^{-\nu\tau}$, $\nu \neq 0$, $|A|^2 + |B|^2 > 0$. For any choice of constants a, a', b_1, b_2, define ϕ_n and ψ_n by R-B transformations (4.40) and (4.42), respectively. Then there exist solutions ϕ_{n+1} [as defined by (4.41)] and ψ_{n+1} [as defined by (4.43)], which are equal, $\psi_{n+1} = \phi_{n+1}$.

To prove this theorem, we define a function ϕ in terms of $(\phi_n, \psi_n, \phi_{n-1})$ and show this function ϕ satisfies both (4.41) and (4.43). Equations (4.40) and (4.42) yield

$$\phi_{n+1,x} = \phi_{n-1,x} + \frac{2a'}{\nu} F[\tfrac{1}{2}(\phi_{n+1} + \phi_n)] + \frac{2a}{\nu} F[\tfrac{1}{2}(\phi_n + \phi_{n-1})], \tag{4.44a}$$

$$\phi_{n+1,y} = \phi_{n-1,y} + \frac{2}{a'}\sinh\left(\frac{\nu}{2}(\phi_{n+1} - \phi_n)\right) - \frac{2}{a}\sinh\left(\frac{\nu}{2}(\phi_n - \phi_{n-1})\right), \tag{4.44b}$$

while (4.42) and (4.43) yield

$$\psi_{n+1,x} = \phi_{n-1,x} + \frac{2a}{\nu} F[\tfrac{1}{2}(\psi_{n+1} + \psi_n)] + \frac{2a'}{\nu} F[\tfrac{1}{2}(\psi_n + \phi_{n-1})], \tag{4.45a}$$

$$\psi_{n+1,y} = \phi_{n-1,y} + \frac{2}{a}\sinh\left(\frac{\nu}{2}(\psi_{n+1} - \psi_n)\right) - \frac{2}{a'}\sinh\left(\frac{\nu}{2}(\psi_n - \phi_{n-1})\right). \tag{4.45b}$$

Clearly, if $\phi_{n+1} = \psi_{n+1} = \phi$, Eqs. (4.44) and (4.45) yield the following two consistency relations:

$$a'F[\tfrac{1}{2}(\phi + \phi_n)] + aF[\tfrac{1}{2}(\phi_n + \phi_{n-1})] = aF[\tfrac{1}{2}(\phi + \psi_n)] + a'F[\tfrac{1}{2}(\psi_n + \phi_{n-1})], \quad (4.46a)$$

$$a \sinh\left(\frac{\nu}{2}(\phi - \phi_n)\right) - a' \sinh\left(\frac{\nu}{2}(\phi_n - \phi_{n-1})\right) = a' \sinh\left(\frac{\nu}{2}(\phi - \psi_n)\right) - a \sinh\left(\frac{\nu}{2}(\psi_n - \phi_{n-1})\right). \quad (4.46b)$$

Using the definition of $F(\cdot)$ and standard hyperbolic identities, we rewrite these consistency conditions as

$$[a \sinh\theta_+ - a' \sinh\theta_-][Ae^{\Sigma} + Be^{-\Sigma}] = 0, \quad (4.47a)$$

$$[a \sinh\theta_+ - a' \sinh\theta_-] \cosh\theta_- = 0, \quad (4.47b)$$

where

$$\theta_\pm \equiv \frac{\nu}{4}[\phi - \phi_{n-1} \pm (\psi_n - \phi_n)],$$

$$\Sigma \equiv \frac{\nu}{4}[\phi + \phi_{n-1} + \psi_n + \phi_n].$$

Thus, the consistency conditions are satisfied if

$$a \sinh\theta_+ = a' \sinh\theta_-. \quad (4.48)$$

We now use Eq. (4.48) to define implicitly a function ϕ in terms of $(\psi_n, \phi_n, \phi_{n-1})$. Taking $\partial/\partial x$ of (4.48), we obtain

$$[a \cosh\theta_+ - a' \cosh\theta_-]\phi_{,x} = \phi_{n,x}[a' \cosh\theta_- + a \cosh\theta_+] + \phi_{n-1,x}[-a' \cosh\theta_- + a \cosh\theta_+] + \psi_{n,x}[-a' \cosh\theta_- - a \cosh\theta_+]. \quad (4.49)$$

Using (4.40a) twice and (4.42a) once, we obtain from (4.49)

$$[a \cosh\theta_+ - a' \cosh\theta_-]\phi_{,x} = \{a \cosh\theta_+ - a' \cosh\theta_-\} \times \left(\phi_{n,x} + \frac{2a'}{\nu}F[\tfrac{1}{2}(\phi + \phi_n)]\right) + \frac{2}{\nu}E, \quad (4.50)$$

where E is given by

$$E = -a'^2 \cosh\theta_-\{F[\tfrac{1}{2}(\psi_n + \phi_{n-1})] - F[\tfrac{1}{2}(\phi + \phi_n]\} + aa' \cosh\theta_-\{2F[\tfrac{1}{2}(\phi_n + \phi_{n-1})] - F[\tfrac{1}{2}(\psi_n + \phi_{n-1})] - F[\tfrac{1}{2}(\phi + \phi_n)]\}. \quad (4.51)$$

E as defined by (4.51) vanishes identically as may be seen by using (4.48) once together with standard hyperbolic identities. Thus ϕ, as defined by (4.48), satisfies (4.41a). Similar arguments establish that it also satisfies (4.41b) and (4.43a, b). This completes the proof of Theorem 3.

Equation (4.48) defining ϕ may be explicitly solved for ϕ. The result of this computation is summarized in the following corollary.

Corollary: Under the hypothesis of Theorem 3,

$$\phi_{n+1} = \psi_{n+1} = \phi,$$

$$\phi = \frac{4}{\nu} \tanh^{-1}\left[\left(\frac{a + a'}{a - a'}\right) \tanh\left(\frac{\nu(\phi_n - \psi_n)}{4}\right)\right] + \phi_{n-1}. \quad (4.52)$$

Theorem (3) is particularly useful when $F(\cdot)$ has a zero. For example, Lamb[13] and Scott[4] use a special case of this theorem to investigate interacting progressive waves (solitons) of the sine-Gordon equation. Selecting $\phi_{n-1} = 0$, the equilibrium solution, they use Bäcklund transformations to generate two solitons,

$$\phi_1 = 4 \tan^{-1}[\exp(ax + y/a + b_1)], \quad \psi_1 = 4 \tan^{-1}[\exp(a'x + y/a' + b_2)]. \quad (4.53)$$

An interaction of these two solitons is then described by ϕ of (4.52) as

$$\phi = 4 \tan^{-1}\left[\left(\frac{a + a'}{a - a'}\right) \times \left(\frac{\exp(ax + y/a + b_1) - \exp(a'x + y/a' + b_2)}{1 + \exp(ax + y/a + b_1)\exp(a'x + y/a' + b_2)}\right)\right]. \quad (4.54)$$

This interaction was obtained by Seeger, Donth, and Kochendörfer[14] in 1953, and later by Perring and Skyrme.[3] The latter authors initially observed the effect numerically. It can represent either a "soliton-soliton" or a "soliton-antisoliton" interaction depending upon the signs of a and a'. Lamb[13] extends his analysis to study more complicated interaction processes by using a sequence of "interlocking" Lamb diagrams.

It is of interest to use these Bäcklund techniques to study interactions governed by more general $F(\cdot)$ than $\sin(\cdot)$. $F(\cdot) = \exp(\cdot)$ and $F(\cdot) = \sinh(\cdot)$ are particularly interesting.

E. Vacuum transformation (P_0, Q_0) and Lamb diagrams

It is of interest to determine whether commuting Lamb diagrams similar to Figs. 3 and 6 can include the vacuum transformation (P_0, Q_0) defined in (4.6). In this section we show that such an assumption is not valid. The vacuum transformations (4.6) do not commute with the general transformations (4.32). To see this, consider the special case $F = \sin\phi$, ϕ real. Then for $\phi_0 = 0$ a commuting diagram similar to Fig. 6 would represent the equations

$$\frac{1}{a_0}\phi_{1,x} = a_0\phi_{1,y} = \sqrt{2(E - \cos\phi_1)}, \quad (4.55a)$$

$$\frac{1}{a_0'}\psi_{1,x} = a_0'\psi_{1,y} = \sqrt{2(E - \cos\psi_1)}, \quad (4.55b)$$

$$\phi_x = 2a_0' \sin\left(\frac{\phi + \phi_1}{2}\right) + \phi_{1,x}, \quad (4.56a)$$

$$= 2a_0 \sin\left(\frac{\phi + \psi_1}{2}\right) + \psi_{1,x}, \quad (4.56b)$$

$$\phi_y = \frac{2}{a_0'} \sin\left(\frac{\phi - \phi_1}{2}\right) - \phi_{1,y}, \quad (4.57a)$$

$$= \frac{2}{a_0} \sin\left(\frac{\phi - \psi_1}{2}\right) - \psi_{1,y}. \quad (4.57b)$$

If the diagram is valid, the value of ϕ obtained by equating the right-hand sides of (4.56a) and (4.56b) must satisfy

$$\phi_{xy} = \sin\phi. \quad (4.58)$$

This equating yields

$$M \sin\frac{\phi}{2} + N \cos\frac{\phi}{2} \stackrel{?}{=} R, \tag{4.59}$$

where

$$M \equiv 2a_0' \cos\frac{\phi_1}{2} - 2a_0 \cos\frac{\psi_1}{2}, \tag{4.60a}$$

$$N \equiv 2a_0' \sin\frac{\phi_1}{2} - 2a_0 \cos\frac{\psi_1}{2}, \tag{4.60b}$$

$$R \equiv a_0'\sqrt{2(E - \cos\psi_1)} - a_0\sqrt{2(E - \cos\phi_1)}. \tag{4.60c}$$

Equation (4.59) is easily solved to yield

$$\sin\frac{\phi}{2} \stackrel{?}{=} (MR \pm N\sqrt{M^2 + (1 - R^2)N^2})/(M^2 + N^2). \tag{4.61}$$

However, (4.61) is useful only if the value of ϕ it determines does indeed satisfy (4.58). Let us suppose $\phi_1, \psi_1 \ll 1$ and calculate (4.61) to first order. Then $M \approx 2(a_0' - a_0), N \approx (a_0'\phi_1 - a_0\psi_1)$, and $R \approx (a_0' - a_0)\sqrt{2(E-1)}$; and (4.61) becomes

$$\sin\frac{\phi}{2} \approx \sqrt{\tfrac{1}{2}(E-1)} \pm (a_0'\phi_1 - a_0\psi_1)\frac{\sqrt{6-2E}}{4(a_0' - a_0)}. \tag{4.62}$$

For $E = 1$ this reduces to

$$\phi \approx \pm (a_0'\phi_1 - a_0\psi_1)/(a_0' - a_0) \tag{4.63}$$

and for $a_0' \gg a_0$, $\phi \approx \phi_1$ as we expect. However, for $E \neq 1$ and $\phi_1 = \psi_1 = 0$, (4.62) says

$$\sin\frac{\phi}{2} \stackrel{?}{=} \sqrt{\tfrac{1}{2}(E-1)} \neq 0 \tag{4.64}$$

which is manifestly incorrect.

Thus a commuting Lamb diagram cannot be used to find the R-B transformation for a periodic solution; direct integration of the general transformations (4.32) must be employed.[14]

F. Comments and speculation

One of the most interesting features of the R-B transformations associated with $\phi_{xy} = F(\phi)$ is the close connection between the transformations and the stable progressive waves (solitons). In Sec. 4D a means for displaying interactions between solitons was investigated. Beyond leading to these explicit interactions, the R-B transformations offer a constructive means to add an additional soliton to a given solution with arbitrarily chosen position and velocity. In principle, one should be able to build a rather large class of solutions by integration of a sequence of first order partial differential equations of type (2.9). Currently we are investigating the size and characterization of this class of solutions. In particular, is it in any sense complete?

Another intriguing feature is the necessary and sufficient condition $F'' = \kappa F$. We have noted that this condition is sufficient to insure the existence of an infinite number of polynomial conservation laws for $\phi_{xy} = F(\phi)$.[32] The relationship between R-B transformations and the existence of an infinite number of conservation laws will be discussed in a second paper. There we consider R-B transformations for the Burgers and Korteweg-deVries equation as well as these nonlinear Klein-Gordon equations.

Finally, we note that the work of Loewner[21] indicates a possible connection between Bäcklund transformations and the inverse method as described by Lax.[16] The Korteweg-deVries equation,[18] and, most recently, the sine-Gordon equation[33] have been solved by this method.

5. THE EQUATION $\phi_{XY} + \alpha\phi_X + \beta\phi_Y = F(\phi)$

In this last section we state a result about the equation $\phi_{xy} + \alpha\phi_x + \beta\phi_y = F(\phi)$, α, β const, $F(\cdot)$ strictly nonlinear; namely, no R-B transformation exists unless $\alpha = \beta = 0$.

Theorem 4: Let $F(\tau)$ be a strictly nonlinear, analytic function of τ. Define the class of R-B transformations to be those solutions $[P(\tau, \xi_1, \xi_2), Q(\tau, \xi_1, \xi_3)]$ of (2.8) which are analytic in each argument $(\tau, \xi_1, \xi_2, \xi_3)$. Further restrict the transformations (P, Q) to depend upon at least one member of the set (ξ_1, ξ_2, ξ_3). If $|\alpha|^2 + |\beta|^2 > 0$, no such transformations (P, Q) exist.

This theorem may be established by a proof very similar to the proof of Theorem 1. Since the proof is quite lengthy, we omit it. We have not removed the exceptional case $F(\cdot)$ a fifth order polynomial through calculations analogous to Appendix B.

Note added in proof: Our current understanding of the relationships between Bäcklund transformations, the inverse method, and conservation laws is presented in Refs. 34 and 35. Additional recent references on the Bäcklund transformation are in Ref. 36.

ACKNOWLEDGMENTS

We wish to express our appreciation to those who organized and participated in the Conference on Nonlinear Wave Propagation sponsored by the American Mathematical Society held at Clarkson College, July 1972; much of this work was begun there. We would also like to acknowledge several helpful conversations with Professor A. Callegari during early stages of this work.

APPENDIX A

Consider Eqs. (4.17a, b) and (4.18a, b) with $p = \bar{p}$ (const) and $q = \bar{q}$ (const). In this appendix we establish the following lemma:

Lemma A1: Under the hypothesis of Theorem 1, $\bar{p} \neq \bar{q}$, $\partial p_0/\partial\tau \not\equiv 0$, $\partial q_0/\partial\tau \not\equiv 0$, $\bar{p} \neq 0$, $\bar{q} \neq 0$, $\partial p_0/\partial\xi_1 \not\equiv 0$, $\partial q_0/\partial\xi_1 \not\equiv 0$.

To prove this lemma, we begin with the assumption that $\bar{p} = \bar{q}$ in (4.17) and (4.18). Under this assumption we subtract (4.18b) from (4.18a) and obtain

$$q_0\frac{\partial p_0}{\partial\tau} = p_0\frac{\partial q_0}{\partial\tau}. \tag{A1}$$

Neither q_0 or p_0 vanishes identically. If $\partial q_0/\partial\tau \equiv \partial p_0/\partial\tau \equiv 0$, then (4.18) yields a contradiction. Therefore, (A1) implies

$$q_0 = A(\xi_1)p_0, \quad A(\xi_1) \not\equiv 0. \tag{A2}$$

Inserting (A2) into (4.17) and using (4.22), we obtain

$$\frac{d}{d\xi_1}A(\xi_1) \equiv 0, \tag{A3}$$

which implies $A = \bar{A}$, A const. Thus, we are left with two equations which must be satisfied.

$$\bar{A}p_0 \frac{\partial p_0}{\partial \tau} + F(\xi_1)\bar{q} - F(\tau) \equiv 0, \tag{A4a}$$

$$\bar{q} \frac{\partial p_0}{\partial \tau} + \frac{\partial p_0}{\partial \xi_1} \equiv 0. \tag{A4b}$$

If $\bar{q} = 0$, then $\partial p_0/\partial \xi_1 \equiv \partial q_0/\partial \xi_1 \equiv 0$, and the solution (P, Q) is independent of (ξ_1, ξ_2, ξ_3). This is contradictory, and $\bar{q} \neq 0$. A similar contradiction follows if $\partial p_0/\partial \xi_1 \equiv 0$. Thus, (A4b) implies $p_0 = \tilde{p}_0(\tau - \bar{q}\xi_1)$. Defining new variables $\tau - \bar{q}\xi_1 \equiv u$, $\tau + \bar{q}\xi_1 \equiv v$ in Eq. (A4a) and taking $\partial/\partial v$ of this equation yields the contradiction $F'(\xi_1) \equiv F'(\tau)$. Thus, (A4a, b) have no solution and $\bar{q} \neq \bar{p}$.

Next we assume $\partial p_0/\partial \tau \equiv 0$. Equation (4.18a) then implies $F(\xi_1)\bar{p} \equiv F(\tau)$, which cannot be. Thus $\partial p_0/\partial \tau \not\equiv 0$; similarly, $\partial q_0/\partial \tau \not\equiv 0$.

Assume now $\bar{p} = 0$. Equation (4.17b) then implies $\partial q_0/\partial \xi_1 \equiv 0$, $q_0 = \bar{q}_0(\tau)$. Since $\bar{q} \neq \bar{p}$, $\bar{q} \neq 0$. If p_0 is independent of ξ_1, Eq. (4.18b) is contradictory. Therefore, (4.17a) implies $p_0 = \tilde{p}_0(\tau - \bar{q}\xi_1)$. This, together with (4.18a) yields

$$\bar{q}_0(\tau)\tilde{p}_0'(\tau - \bar{q}\xi_1) = F(\tau) \Rightarrow \tilde{p}_0(\tau - \bar{q}\xi_1) = \kappa_1(\tau - \bar{q}\xi_1) + \kappa_2, \quad \kappa_1 \text{ and } \kappa_2 \text{ const.} \tag{A5}$$

Inserting (A5) back into (4.18a) implies $\kappa_1 \neq 0$ and $\bar{q}_0(\tau) = F(\tau)/\kappa_1$. Inserting these expressions for $\bar{q}_0$ and $\tilde{p}_0$ into (4.18b) yields the contradiction $F'(\tau) \equiv F'(\xi_1)$. Thus $\bar{p} \neq 0$; similarly, $\bar{q} \neq 0$.

Finally assume $\partial p_0/\partial \xi_1 \equiv 0$. Then Eq. (4.17a) implies $\partial p_0/\partial \tau \equiv 0$. This fact, together with (4.18a) yields the contradiction $F(\xi_1)\bar{p} \equiv F(\tau)$. Thus $\partial p_0/\partial \tau \not\equiv 0$; similarly, $\partial q_0/\partial \tau \not\equiv 0$. This completes the proof of Lemma A1.

APPENDIX B

In this appendix we prove the following lemma:

Lemma B1: Consider $F(\tau) = \Gamma + \gamma_1\tau + \gamma_2\tau^2$, $\gamma_2 \neq 0$. Under the hypothesis of Theorem 1, no R-B transformation (P, Q) exists.

We begin the proof of this lemma by noticing that condition (4.29) yields

$$2\gamma_2^2(1 - 1/c)\tau^2 - 2\gamma_2^2(1 - c)\xi_1^2 + 2\gamma_1\gamma_2(1 - 1/c)\tau - 2\gamma_1\gamma_2(1 - c)\xi_1 + 2\gamma_2\Gamma(c - 1/c) \equiv 0. \tag{B1}$$

Since $\gamma_2 \neq 0$, (B1) implies $c = +1$. With $c = +1$ and this quadratic $F(\cdot)$, (4.26a, b) become

$$\bar{q}_0(u)\tilde{p}_0'(v) \equiv (\gamma_2/2)(u^2 + v^2) + \gamma_1 u + 2\Gamma, \tag{B2a}$$

$$\tilde{p}_0(v)\bar{q}_0'(u) \equiv \gamma_2 uv + \gamma_1 v. \tag{B2b}$$

Solving (B2b) for $\tilde{p}_0(v)$ and calculating $(\partial/\partial v)\tilde{p}_0(v)$ yields

$$\tilde{q}_0(u)\tilde{p}_0'(v) = [\bar{q}_0(u)/\bar{q}_0'(u)][\gamma_2 u + \gamma_1]. \tag{B3}$$

But $\bar{q}_0(u)\tilde{p}_0'(v)$ is also given by (B2a) which, together with (B3), yields

$$\frac{\bar{q}_0(u)}{\bar{q}_0'(u)} = \frac{(\gamma_2/2)(u^2 + v^2) + \gamma_1 u + 2\Gamma}{\gamma_2 u + \gamma_1}. \tag{B4}$$

Equation (B4) contradicts the assumption $\gamma_2 \neq 0$. The proof is complete.

APPENDIX C

In this appendix we establish that the constant κ_1 in (4.31) is not equal to zero.

Integrating (4.31) once yields

$$\tfrac{1}{2}(F'(\tau))^2 = \tfrac{1}{2}\kappa F^2(\tau) + \kappa_1 F(\tau) + \kappa_2, \quad \kappa_2 \text{ const.} \tag{C1}$$

Inserting (4.31) and (C1) into (4.29), we obtain

$$0 \equiv [F(\tau) - F(\xi_1)](1 - c), \tag{C2}$$

which implies $c = +1$.

Finally, if $F(\cdot)$ satisfies (4.31), it is of the form

$$F(\tau) = Ae^{\nu\tau} + Be^{-\nu\tau} - \kappa_1/\kappa, \quad \text{where } \nu^2 = \kappa \text{ and } |A|^2 + |B|^2 > 0. \tag{C3}$$

Inserting (C3) into (4.26) and using the fact $c = +1$, we obtain

$$\bar{q}_0(u)\tilde{p}_0'(v) \equiv 2\cosh(\nu v/2)(Ae^{(\nu/2)u} + Be^{-(\nu/2)u}) - 2\kappa_1/\kappa, \tag{C4a}$$

$$\tilde{p}_0(v)\bar{q}_0'(u) \equiv 2\sinh(\nu v/2)(Ae^{(\nu/2)u} - Be^{-(\nu/2)u}), \tag{C4b}$$

which integrate to yield

$$\bar{q}_0(u)\tilde{p}_0(v) \equiv (4/\nu)\sinh(\nu v/2)(Ae^{(\nu/2)u} + Be^{-(\nu/2)u}) - \frac{2\kappa_1}{\kappa}v + \kappa_3(u), \tag{C5a}$$

$$\bar{q}_0(u)\tilde{p}_0(v) \equiv (4/\nu)\sinh(\nu v/2)(Ae^{(\nu/2)u} + Be^{-(\nu/2)u}) + \kappa_4(v), \tag{C5b}$$

where "constants" of integration clearly must be related by $\kappa_4 = \kappa_3 - (2\kappa_1/\kappa)v$, κ_3 const. Equation (C4b) implies $\tilde{p}_0(0) = 0$, which together with (C5b) implies $\kappa_3 = 0$. Solving (C4a) for $q_0(u)$, multiplying by $\tilde{p}_0(v)$, and equating to (C5a) yields

$$[\tilde{p}_0(v)/\tilde{p}_0'(0)](Ae^{(\nu/2)u} + Be^{-(\nu/2)u} - \kappa_1/\kappa) \equiv (2/\nu)\sinh(\nu v/2)(Ae^{(\nu/2)u} + Be^{-(\nu/2)u}) - (2\kappa_1/\kappa)v. \tag{C6}$$

Calculating $\partial/\partial u$ of (C6), we find

$$\tilde{p}_0(v)/\tilde{p}_0'(0) \equiv (2/\nu)\sinh(\nu v/2), \tag{C7}$$

which, when inserted back into (C5) yields

$$\sinh(\nu v/2)(Ae^{(\nu/2)u} + Be^{-(\nu/2)u} - \kappa_1/\kappa) = \sinh(\nu v/2)(Ae^{(\nu/2)u} + Be^{-(\nu/2)u}) - (\kappa_1/\nu)v. \tag{C8}$$

Equation (C8) contradicts the assumption $\kappa_1 \neq 0$.

*This work has been supported in part by the National Science Foundation under Grants No. GK-12502 and GP-37627.

[1]N. J. Zabusky and C. J. Galvin, J. Fluid Mech. **47**, 811 (1971).
[2]C. H. Su and C. S. Gardner, J. Math. Phys. **10**, 536 (1969).
[3]J. K. Perring and T. H. R. Skyrme, Nucl. Phys. **31**, 550 (1962).
[4]A. C. Scott, Nuovo Cimento B **69**, 241 (1970).

[5]M. Toda and M. Wadati, J. Phys. Soc. Jap. **34,** 18 (1973).
[6]G. L. Lamb, Rev. Mod. Phys. **43**, 99 (1971).
[7]N. J. Zabusky and M. D. Kruskal, Phys. Rev. Lett. **15**, 240 (1965).
[8]F. Tappert, lecture given at AMS Summer Seminar on Nonlinear Wave Motion, Clarkson College of Technology, Potsdam, New York (July, 1972).
[9]N. J. Zabusky, in *Nonlinear Partial Differential Equations*, edited by W. F. Ames (Academic, New York, 1967), p. 242.
[10]A. Barone, F. Esposito, C. J. Magee, and A. C. Scott, Riv. Nuovo Cimento **1,** 227 (1971).
[11]W. F. Ames, *Nonlinear Partial Differential Equations in Engineering* (Academic, New York, to appear), Vol. II.
[12]J. Rubinstein, J. Math. Phys. **11,** 258 (1970).
[13]G. L. Lamb, Jr. *In Honor of Philip M. Morse*, edited by H. Feshback and K. U. Ingard (MIT, Cambridge, Mass., 1969), p. 88.
[14]A. Seeger, H. Donth, and A. Kochendörfer, Z. Phys. **34**, 173 (1953).
[15]C. S. Gardner, J. M. Greene, M. D. Kruskal, and R. M. Miura, Phys. Rev. Lett. **19,** 1095 (1967).
[16]P. D. Lax, Commun. Pure Appl. Math. **21,** 467 (1968).
[17]R. M. Miura, C. S. Gardner, and M. D. Kruskal, J. Math. Phys. **9,** 1204 (1968).
[18]V. E. Zakharov and A. B. Shabat, Zh. Eksp. Teor. Fiz. **61,** 118 (1971) [Sov. Phys.-JETP **34,** 62 (1972)].
[19]G. B. Whitham, private communication.
[20]C. Loewner, NACA, Technical Note 2065 (1965).
[21]C. Loewner, J. Analyse Math. **2,** 219 (1953).
[22]C. Rogers, "Transformari Bäcklund in magnetogazodinamica Liniara nedisipativa," presented to Conferinta Nationala de Mercanica Aplicata, Romania (1969).
[23]C. Rogers, Acta Phys. Austriaca **31,** 80 (1970).
[24]C. Rogers, Ann. Soc. Sci. Brux. I **86,** 211 (1972).
[25]G. Power, C. Rogers, and R. A. Osbourne, Z. Angew. Math. Mech. **49,** 333 (1969).
[26]A. V. Bäcklund, Math. Ann. **9,** 297 (1876).
[27]A. V. Bäcklund, Math. Ann. **19,** 387 (1882).
[28]M. J. Clairin, Ann. Fac. Sci. Univ. Toulouse **5,** 437 (1903).
[29]E. Goursat, Ann. Fac. Sci. Univ. Toulouse
[30]A. R. Forsyth, *Theory of Differential Equations* (Dover, New York, 1959), Vol. 6, Chap. 21.
[31]L. P. Eisenhart, *A Trieste on the Differential Geometry of Curves and Surfaces* (Dover, New York, 1960).
[32]M. D. Kruskal, lecture given at AMS Summer Seminar on Nonlinear Wave Motion, Clarkson College of Technology, Potsdam, New York (July, 1972).
[33]G. L. Lamb, Jr., Bull. Am. Phys. Soc. **18,** 64 (1973).
[34]A. C. Scott, F. Y. F. Chu and D. W. McLaughlin, Proc. IEEE **61** (1973).
[35]G. L. Lamb, Jr. (to be published).
[36]M. Matsuda, Trans. Am. Math. Soc. **150**, 327 (1970); **166**, 371 (1972).

Kinetic equations for the quantized motion of a particle in a randomly perturbed potential field*†

Ioannis M. Besieris

Department of Electrical Engineering, Virginia Polytechnic Institute and State University, Blacksburg, Virginia 24061

Frederick D. Tappert

Bell Laboratories, Whippany, New Jersey 07981

(Received 6 September 1972)

Within the framework of the first-order smoothing approximation and the long-time, Markovian approximation, kinetic equations are derived for the stochastic Wigner equation (the exact equation of evolution of the phase-space Wigner distribution function) and the stochastic Liouville equation (correspondence limit approximation) associated with the quantized motion of a particle described by a stochastic Schrödinger equation. In the limit of weak fluctuations and long times, the transport equation for the average probability density of the particle in momentum space which was reported recently by Papanicolaou is recovered. Also, on the basis of the Novikov functional formalism, it is established that several of the approximate kinetic equations derived in this paper are identical to the exact statistical equations in the special case that the potential field is a δ-correlated (in time), homogeneous, wide-sense stationary, Gaussian process.

1. INTRODUCTION

Consider a particle whose wave function evolves according to the stochastic Schrödinger equation:

$$i\hbar\frac{\partial}{\partial t}\psi(\underline{x},t;\alpha)=H_{\rm op}\left(\underline{x},-i\hbar\frac{\partial}{\partial\underline{x}},t;\alpha\right)\psi(\underline{x},t;\alpha),\qquad t\geq 0,\quad \underline{x}\in R^3, \tag{1.1a}$$

$$H_{\rm op}\left(\underline{x},-i\hbar\frac{\partial}{\partial\underline{x}},t;\alpha\right)=-\frac{\hbar^2}{2m}\nabla^2+V(\underline{x},t;\alpha). \tag{1.1b}$$

Here, ∇^2 denotes the Laplacian in R^3 and $H_{\rm op}$ is a self-adjoint, stochastic operator depending on a parameter $\alpha\in\mathfrak{A}$, $\mathfrak{A}$ being a probability measure space. In addition, $\psi(\underline{x},t;\alpha)$, the complex random wave function, is an element of an infinitely dimensional vector space $\mathfrak{K}$, $\hbar$ is Planck's constant divided by 2π, m is the mass of the particle, and V is the potential field which is assumed to be a real, space- and time-dependent random function. Equation (1.1) is rendered closed by specifying Cauchy initial data, viz., $\psi(\underline{x},0;\alpha)=\psi_0(\underline{x};\alpha)$.

It is our intent in this paper to derive kinetic or transport equations for the average phase-space "probability" density of a quantized particle whose motion is described by the stochastic Schrödinger Eq. (1.1). Knowledge of this quantity enables one to obtain such physically important averaged entities as the probability density in configuration and momentum space, the probability current density, the angular momentum, as well as the centroid and width of a wave packet. In this respect, we extend the work of previous workers[1] who have considered the stochastic motion of a particle in the classical or correspondence limit ($\hbar\to 0$) approximation, either abstractly or in connection with physical applications, such as acceleration of cosmic rays, heating of thermonuclear plasma, turbulence of interstellar plasma, etc. The random quantum-mechanical harmonic oscillator problem has also been treated by several workers.[2]

In Sec. 2, stochastic equations describing the evolution of the phase-space Wigner distribution function associated with (1.1) are given in various useful representations. In order for the discussion to be self-contained, a brief exposition is provided in Sec. 3 of the first-order smoothing approximation and the long-time, Markovian approximation which constitute the basis for the statistical analysis in this paper. In Sec. 4, a kinetic equation corresponding to the stochastic Wigner equation is derived within the range of applicability of the first-order smoothing approximation. In the long-time Markovian approximation, we obtain a novel transport equation for the ensemble-averaged Wigner distribution function which has the classical form of a radiation transport equation, or a Boltzmann equation for waves (quasiparticles in phase space). Integration of this equation over $\underline{x}$-space gives a transport equation for the average probability density of the particle in momentum space, which coincides with the result reported recently by Papanicolaou.[3] Four special cases of the kinetic equation are derived in Sec. 5 and relationships with radiation transport equations are discussed. In Sec. 6 we consider the kinetic theory of the stochastic Liouville equation and derive Fokker–Planck equations which are of importance in practical applications of the theory. Finally, in the Appendix it is shown by means of the Novikov functional formalism (cf. Ref. 4) that for a δ-correlated (in time), homogeneous, wide-sense stationary, Gaussian process, one can derive exact transport equations for both the stochastic Wigner equation and the stochastic Liouville equation. It is interesting to note that these exact kinetic equations are identical with the corresponding ones derived in Secs. 5 and 6 on the basis of the first-order smoothing approximation, without, however, the restriction that the random process be Gaussian.

2. THE WIGNER DISTRIBUTION FUNCTION

In the following we shall be concerned with the time evolution of a "measurable" quantity. In this respect, the wave function $\psi(\underline{x},t;\alpha)$ has little physical meaning. We may, however, consider the total wave energy and the total wave action (probability) which are given in terms of the wave function ψ and the operator $H_{\rm op}$ by the integrals

$$E=\int_{R^3}\psi^*H_{\rm op}\psi\,d\underline{x}, \tag{2.1}$$

$$A=\int_{R^3}\psi^*\psi\,d\underline{x}. \tag{2.2}$$

The total wave energy is not conserved since the poten-

tial is taken to be time-dependent. On the other hand, the total wave action is conserved by virtue of the self-adjointness of H_{op}. The integrands in (2.1) and (2.2) are respectively the space wave energy density function and the space wave action density function (probability density function in configuration space).

The equal-time, two-point density function for a pure state is introduced next as follows in terms of the wave function:

$$\rho(\underline{x}_2, \underline{x}_1, t; \alpha) = \psi^*(\underline{x}_2, t; \alpha)\psi(\underline{x}_1, t; \alpha). \tag{2.3}$$

It obeys the von Neumann equation

$$i\hbar\frac{\partial}{\partial t}\rho(\underline{x}_2, \underline{x}_1, t; \alpha) = \left(-\frac{\hbar^2}{2m}\nabla^2_{\underline{x}_1} + \frac{\hbar^2}{2m}\nabla^2_{\underline{x}_2} + V(\underline{x}_1, t; \alpha) - V(\underline{x}_2, t; \alpha)\right) \times \rho(\underline{x}_2, \underline{x}_1, t; \alpha). \tag{2.4}$$

The phase-space analog of the density function is provided by the Wigner distribution function which is defined as follows:[5]

$$f(\underline{x}, \underline{p}, t; \alpha) = (2\pi\hbar)^{-3}\int_{R^3} d\underline{y} e^{i\underline{p}\cdot\underline{y}/\hbar}\ \rho(\underline{x} + \tfrac{1}{2}\underline{y}, \underline{x} - \tfrac{1}{2}\underline{y}, t; \alpha). \tag{2.5}$$

This quantity is real, but not necessarily positive everywhere. The total wave energy and wave action can be written in terms of the Wigner distribution function as follows:

$$E = \int_{R^3} d\underline{x}\int_{R^3} d\underline{p} H(\underline{x}, \underline{p}, t; \alpha) f(\underline{x}, \underline{p}, t; \alpha), \tag{2.6a}$$

$$A = \int_{R^3} d\underline{x}\int_{R^3} d\underline{p} f(\underline{x}, \underline{p}, t; \alpha). \tag{2.6b}$$

Here, $H(\underline{x}, \underline{p}, t; \alpha)$ is the Weyl transform of the operator H_{op} and is given explicitly as

$$H(\underline{x}, \underline{p}, t; \alpha) = \frac{1}{2m}p^2 + V(\underline{x}, t; \alpha), \qquad p = |\underline{p}|. \tag{2.7}$$

Examining (2.6) reveals that $H(\underline{x}, \underline{p}, t; \alpha)f(\underline{x}, \underline{p}, t; \alpha)$ can be interpreted as the phase-space energy density. Its integral over $\underline{p}$-space provides the space energy density. Similarly, $f(\underline{x}, \underline{p}, t; \alpha)$ can be thought of as the phase-space "probability" density. Its integral over $\underline{p}$-space is the *correct* probability density in configuration space. Also, its integral over $\underline{x}$-space is the correct probability density in momentum space. Finally, the integral of $(\underline{p}/m)f(\underline{x}, \underline{p}, t; \alpha)$ over $\underline{p}$-space is recognized as the conventional expression for the probability current density in configuration space.

Using the definition of $f(\underline{x}, \underline{p}, t; \alpha)$ given in (2.5) in conjunction with the von Neumann equation (2.4), it is found that the Wigner distribution function evolves according to the equation

$$\frac{\partial}{\partial t}f(\underline{x}, \underline{p}, t; \alpha) = \mathfrak{L}f(\underline{x}, \underline{p}, t; \alpha), \tag{2.8a}$$

$$\mathfrak{L}f(\underline{x}, \underline{p}, t; \alpha) = -\frac{1}{m}\underline{p}\cdot\frac{\partial}{\partial \underline{x}}f(\underline{x}, \underline{p}, t; \alpha) + \theta f(\underline{x}, \underline{p}, t; \alpha). \tag{2.8b}$$

On the basis of the work of Wigner (cf. Ref. 5), Groenewold,[6] Moyal,[7] Irving and Zwanzig,[8] and the recent findings of Leaf,[9] the following representations of the potential-dependent term on the right-hand side of (2.8b) are possible:

(i) $$\theta f(\underline{x}, \underline{p}, t; \alpha) = \int_{R^3} d\underline{p}' K(\underline{x}, \underline{p} - \underline{p}', t; \alpha) f(\underline{x}, \underline{p}', t; \alpha),$$
$$K(\underline{x}, \underline{p}, t; \alpha) = (i\hbar)^{-1}(2\pi\hbar)^{-3}\int_{R^3} d\underline{y} e^{i\underline{p}\cdot\underline{y}/\hbar} \times [V(\underline{x} - \tfrac{1}{2}\underline{y}, t; \alpha) - V(\underline{x} + \tfrac{1}{2}\underline{y}, t; \alpha)]; \tag{2.9a}$$

(ii) $$\theta f(\underline{x}, \underline{p}, t; \alpha) = (i\hbar)^{-1}(2\pi\hbar)^{-3}\int_{R^3} d\underline{y} e^{i\underline{p}\cdot\underline{y}/\hbar} \times \rho(\underline{x} + \tfrac{1}{2}\underline{y}, \underline{x} - \tfrac{1}{2}\underline{y}, t; \alpha) \times [V(\underline{x} - \tfrac{1}{2}\underline{y}, t; \alpha) - V(\underline{x} + \tfrac{1}{2}\underline{y}, t; \alpha)], \tag{2.9b}$$

(iii) $$\theta f(\underline{x}, \underline{p}, t; \alpha) = V(\underline{x}, t; \alpha)\frac{2}{\hbar}\sin\left(\frac{\hbar}{2}\left(\frac{\overleftarrow{\partial}}{\partial \underline{x}}\cdot\frac{\overrightarrow{\partial}}{\partial \underline{p}}\right)\right) f(\underline{x}, \underline{p}, t; \alpha). \tag{2.9c}$$

These different forms will prove useful in our subsequent work. We shall refer to the exact equation of evolution of $f(\underline{x}, \underline{p}, t; \alpha)$ [cf. Eq. (2.8)], with $\theta f(\underline{x}, \underline{p}, t; \alpha)$ given by any of the representations (2.9), as the *stochastic Wigner equation*.

It is seen from (2.9c) that in the correspondence limit ($\hbar \to 0$),

$$\theta f(\underline{x}, \underline{p}, t; \alpha) = \frac{\partial}{\partial \underline{x}}V(\underline{x}, t; \alpha)\cdot\frac{\partial}{\partial \underline{p}}f(\underline{x}, \underline{p}, t; \alpha) + O(\hbar^2). \tag{2.10}$$

In the framework of this approximation, we shall refer to (2.8) as the *stochastic Liouville equation* satisfied by the Wigner distribution function.

3. GENERAL EQUATIONS FOR THE MEAN WIGNER DISTRIBUTION FUNCTION

We shall consider in this section a statistical analysis of the abstract equation (2.8).

The Wigner distribution function and the operator $\mathfrak{L}$ are separated into mean and fluctuating parts:

$$f(\underline{x}, \underline{p}, t; \alpha) = \mathcal{E}\{f(\underline{x}, \underline{p}, t; \alpha)\} + \delta f(\underline{x}, \underline{p}, t; \alpha), \tag{3.1a}$$

$$\mathfrak{L} = \mathcal{E}\{\mathfrak{L}\} + \delta\mathfrak{L}. \tag{3.1b}$$

Using only the first-order smoothing approximation (cf., e.g., Refs. 10, 11, 12) one obtains a kinetic equation for the ensemble average of the Wigner distribution function:

$$\left(\frac{\partial}{\partial t} - \mathcal{E}\{\mathfrak{L}\}\right)\mathcal{E}\{f(\underline{x}, \underline{p}, t; \alpha)\} = \int_0^t d\tau \mathcal{E}\{\delta\mathfrak{L}(t)\exp[\tau\mathcal{E}\{\mathfrak{L}\}]\delta\mathfrak{L}(t-\tau)\}\mathcal{E}\{f(\underline{x}, \underline{p}, t-\tau; \alpha)\}. \tag{3.2}$$

In deriving (3.2) it has been assumed that the fluctuating part of the initial Wigner distribution function is zero and that $\mathcal{E}\{\mathfrak{L}\}$ does not depend on time. By analogy to a similar equation in quantum electrodynamics, (3.2) is called the *Dyson–Schwinger equation* with

$\int_0^t d\tau \mathcal{E}\{\delta\mathfrak{L}(t)\exp[\tau\mathcal{E}\{\mathfrak{L}\}]\delta\mathfrak{L}(t-\tau)\}\mathcal{E}\{\cdot\}$ *the mass operator*.

This kinetic equation is uniformly valid in time, and applies for a potential field with inhomogeneous deterministic background. The right-hand side of (3.2) contains generalized operators (nonlocal, with memory) in phase-space.

By imposing additional restrictions, various levels of simplification may be obtained. The long-time, Markovian approximation (cf., e.g., Refs. 13, 14) yields the simpler equation

$$\left(\frac{\partial}{\partial t} - \mathcal{E}\{\mathfrak{L}\}\right)\mathcal{E}\{f(\underline{x}, \underline{p}, t; \alpha)\} = \left[\int_0^\infty d\tau \mathcal{E}\{\delta\mathfrak{L}(t)\ \exp[\tau\mathcal{E}\{\mathfrak{L}\}]\delta\mathfrak{L}(t-\tau)\}\right]\mathcal{E}\{f(\underline{x}, \underline{p}, t; \alpha)\}. \tag{3.3}$$

The boldface square brackets on the right-hand side signify that $\mathcal{E}\{f(\underline{x},\underline{p},t;\alpha)\}$ lies outside the influence of the operator $\exp[\tau\mathcal{E}\{\mathfrak{L}\}]$.

Having established an expression for the mean Wigner distribution function by solving either of the above kinetic equations and assuming that the total action (probability) is normalized to unity, the following physically meaningful averaged quantities can be determined by straightforward integration: (i) probability density in configuration space: $\mathcal{E}\{\rho(\underline{x},\underline{x},t;\alpha)\} = \int d\underline{p}\,\mathcal{E}\{f(\underline{x},\underline{p},t;\alpha)\}$; (ii) probability density in momentum space: $\mathcal{E}\{\hat{\rho}(\underline{p},\underline{p},t;\alpha)\} = \int d\underline{x}\,\mathcal{E}\{f(\underline{x},\underline{p},t;\alpha)\}$, where $\hat{\rho}(\underline{p},\underline{p},t;\alpha)$ is the momentum representation of the density function; (iii) probability current density in configuration space: $\mathcal{E}\{\underline{J}\} = \int d\underline{p}(\underline{p}/m)\mathcal{E}\{f(\underline{x},p,t;\alpha)\}$; (iv) centroid of a wave packet: $\underline{x}_c = \int d\underline{x}\int d\underline{p}\underline{x}\,\mathcal{E}\{f(\underline{x},\underline{p},t;\alpha)\}$; (v) spread of a wave packet: $(1/2)\sigma^2 = \int d\underline{x}\int d\underline{p}(\underline{x}-\underline{x}_c)^2\mathcal{E}\{f(\underline{x},\underline{p},t;\alpha)\}$; (vi) angular momentum about a point $\underline{a}$: $\mathcal{E}\{\underline{\mathfrak{L}}\} = \int d\underline{x}\int d\underline{p}[(\underline{x}-\underline{a})\times\underline{p}]\mathcal{E}\{f(\underline{x},\underline{p},t;\alpha)\}$.

4. KINETIC THEORY FOR THE STOCHASTIC WIGNER EQUATION

We shall specialize here the results of the previous section to the stochastic Wigner equation (2.8). It is convenient to use for this purpose the representation (2.9a) for $\theta f(\underline{x},\underline{p},t;\alpha)$.

On the basis of the first-order smoothing approximation, one finds, corresponding to (3.2), the kinetic equation

$$\left(\frac{\partial}{\partial t}+\frac{1}{m}\underline{p}\cdot\frac{\partial}{\partial\underline{x}}\right)\mathcal{E}\{f(\underline{x},\underline{p},t;\alpha)\} = \Theta\,\mathcal{E}\{f(\underline{x},\underline{p},t;\alpha)\}, \tag{4.1a}$$

$$\Theta\mathcal{E}\{f(\underline{x},\underline{p},t;\alpha)\} = \int_0^t d\tau\int_{R^3}d\underline{p}'\int_{R^3}d\underline{p}''\,\mathcal{E}\Big\{\delta K(\underline{x},\underline{p}-\underline{p}',t;\alpha) \times \delta K\Big(\underline{x}-\frac{\tau}{m}\underline{p}',\underline{p}'-\underline{p}'',t-\tau;\alpha\Big)\Big\} \times \mathcal{E}\Big\{f\Big(\underline{x}-\frac{\tau}{m}\underline{p}',\underline{p}'',t-\tau;\alpha\Big)\Big\}. \tag{4.1b}$$

In writing down (4.1) we have resorted to the simplifying assumption that the deterministic background potential field is independent of the space and time coordinates. By virtue of (2.9a), δK, the fluctuating part of K, is related to the fluctuating part of the potential field. For the sake of simplicity, the latter is taken to be spatially homogeneous and wide-sense stationary, viz.,

$$\Gamma(\underline{y},\tau) = \mathcal{E}\{\delta V(\underline{x},t;\alpha)\delta V(\underline{x}-\underline{y},t-\tau;\alpha)\}. \tag{4.2}$$

The correlation function is even in both $\underline{y}$ and τ. The calculations to follow are simplified if we introduce the space-time Fourier transforms

$$\delta V(\underline{x},t;\alpha) = \int_{R^3}d\underline{p}\int_{R^1}du e^{i(\underline{p}\cdot\underline{x}-ut)/\hbar}\,\delta\hat{V}(\underline{p},u;\alpha), \tag{4.3a}$$

$$\Gamma(\underline{y},\tau) = \int_{R^3}d\underline{p}\int_{R^1}du e^{i(\underline{p}\cdot\underline{y}-u\tau)/\hbar}\,\hat{\Gamma}(\underline{p},u). \tag{4.3b}$$

Because δV is a stationary random process, $\delta\hat{V}$ must be regarded as a generalized function. It follows that the Fourier-space analog of (4.2) may be written as

$$\mathcal{E}\{\delta\hat{V}(\underline{p},u)\delta\hat{V}(\underline{p}',u')\} = \delta(\underline{p}+\underline{p}')\delta(u+u')\hat{\Gamma}(\underline{p},u). \tag{4.4}$$

We note that $\hat{\Gamma}$ is real, nonnegative, and even in both arguments.

The operator Θ may now be evaluated by a straightforward calculation. We obtain

$$\Theta\mathcal{E}\{f(\underline{x},\underline{p},t;\alpha)\} = \frac{2}{\hbar^2}\int_{R^3}d\underline{p}'\int_0^t d\tau Q(\underline{p},\underline{p}',\tau) \times\left[\mathcal{E}\Big\{f\Big(\underline{x}-\frac{\tau}{2m}(\underline{p}+\underline{p}'),\underline{p}',t-\tau;\alpha\Big)\Big\} - \mathcal{E}\Big\{f\Big(\underline{x}-\frac{\tau}{2m}(\underline{p}+\underline{p}'),\underline{p},t-\tau;\alpha\Big)\Big\}\right], \tag{4.5a}$$

$$Q(\underline{p},\underline{p}',\tau) = \frac{1}{(2\pi\hbar)^3}\int_{R^3}d\underline{y}\,\Gamma(\underline{y},\tau) \times\cos\left[\underline{y}\cdot(\underline{p}'-\underline{p})/\hbar+\tau\left(\frac{p^2}{2m}-\frac{p'^2}{2m}\right)/\hbar\right]. \tag{4.5b}$$

Introducing (4.5) into (4.1a) yields the phase-space kinetic equation for the mean Wigner distribution function within the limits of the first-order smoothing approximation. This kinetic equation is uniformly valid in time, and no assumptions have been made concerning the scale lengths of the potential fluctuations. The main assumption in the first-order smoothing approximation is that the potential fluctuations be sufficiently small.

In the long-time Markovian approximation, the kinetic equation simplifies considerably. When the limiting process proceeding from (3.2) to (3.3) is applied to Eq. (4.5a), we obtain the kinetic equation:

$$\left(\frac{\partial}{\partial t}+\frac{1}{m}\underline{p}\cdot\frac{\partial}{\partial\underline{x}}\right)\mathcal{E}\{f(\underline{x},\underline{p},t;\alpha)\} = \int_{R^3}d\underline{p}'W(\underline{p},\underline{p}') \times[\mathcal{E}\{f(\underline{x},\underline{p}',t;\alpha)\}-\mathcal{E}\{f(\underline{x},\underline{p},t;\alpha)\}], \tag{4.6}$$

where the transition probability W is given by

$$W(\underline{p},\underline{p}') = \frac{2\pi}{\hbar}\hat{\Gamma}\left(\underline{p}-\underline{p}',\frac{p^2}{2m}-\frac{p'^2}{2m}\right). \tag{4.7}$$

Equation (4.6) has the form of a radiation transport equation, or a Boltzmann equation for waves (quasiparticles in phase space). As expected, the expression for the transition probability, Eq. (4.7), has the form given by the "golden rule" of quantum perturbation theory.[15] From what was said above about $\hat{\Gamma}$, it follows that W is real, nonnegative, and $W(\underline{p},\underline{p}') = W(\underline{p}',\underline{p})$. The latter relation (principle of detailed balance) implies conservation of probability: $\int_{R^3}d\underline{x}\int_{R^3}d\underline{p}\mathcal{E}\{f(\underline{x},\underline{p},t;\alpha)\} =$ const. One must bear in mind, however, that the Wigner distribution function f, as well as $\mathcal{E}\{f\}$, is not in general nonnegative so that it cannot be regarded as a probability density in the usual sense.

We believe that this kinetic equation, as well as the method used to derive it is essentially new. It is the main result of this article. Unlike other derivations of radiation transport equations, the derivation given here does not require that wavelengths be small compared to correlation lengths. The main additional assumption in proceeding to the long-time Markovian approximation is that $\mathcal{E}\{f(\underline{x},\underline{p},t;\alpha)\}$ should vary slowly in $\underline{x}$ and t compared to the correlation lengths and times. But because of the manner in which the Wigner distribution function is constructed, f (and *a fortiori* $\mathcal{E}\{f\}$) varies slowly compared to the wavelengths and periods of the ψ-waves themselves. Of course, if situations arise where the conditions for the validity of the long-time Markovian approximation are not satisfied, one could use the more general kinetic equation given by (4.5) and (4.1).

If the mean Wigner distribution function is normalized to unity, i.e.,

$$\int_{R^3}d\underline{x}\int_{R^3}d\underline{p}\,\mathcal{E}\{f(\underline{x},\underline{p},t;\alpha)\} = 1,$$

then it is well known that

$$D(t) \equiv (2\pi\hbar)^3 \int_{R^3} d\underline{x} \int_{R^3} d\underline{p}[\mathcal{E}\{f(\underline{x},\underline{p},t;\alpha)\}]^2 \leq 1. \quad (4.8)$$

Equality holds if and only if $\mathcal{E}\{f\}$ is a "pure" state. Otherwise $\mathcal{E}\{f\}$ is said to represent a "mixed" state, and D (which we shall call the *degree of coherence*) is less than unity. Assume now that $\mathcal{E}\{f\}$ evolves according to the kinetic equation (4.6). Using the principle of detailed balance, an easy calculation shows that

$$\frac{dD}{dt} = -(2\pi\hbar)^3 \int_{R^3} d\underline{x} \int_{R^3} d\underline{p} \int_{R^3} d\underline{p}' W(\underline{p},\underline{p}') \times [\mathcal{E}\{f(\underline{x},\underline{p},t;\alpha)\} - \mathcal{E}\{f(\underline{x},\underline{p}',t;\alpha)\}]^2 \leq 0. \quad (4.9)$$

Hence the degree of coherence is a monotonically decreasing function of time. Any equilibrium solution of Eq. (4.6) must satisfy $dD/dt = 0$. From (4.9) it is seen that in regions where W is nonzero (so that each state is connected to each other state), the equilibrium solution must be independent of $\underline{p}$. In particular, if W nowhere vanishes, then the only equilibrium solution is the equipartition solution, $\mathcal{E}\{f\} = \text{const}$.

This solution is not normalized, however, and can only be approached asymptotically in time through a sequence of normalized solutions. In Sec. 5, Case (ii), we shall show that if the fluctuations are time-independent, then $W(\underline{p},\underline{p}')$ is proportional to $\delta[(p^2/2m) - (p'^2/2m)]$. In this case, the transition probability connects only states with equal energy and Eq. (4.9) shows that the equilibrium solutions are uniformly distributed on the energy surfaces $p^2/2m = \text{const}$.

A kinetic equation in momentum space, or "master" equation, may be derived by integrating (4.6) over $\underline{x}$-space and recalling that $\mathcal{E}\{\hat{\rho}(\underline{p},\underline{p},t;\alpha)\} = \int_{R^3} d\underline{x}\, \mathcal{E}\{f(\underline{x},\underline{p},t;\alpha)\}$ represents the mean probability density of a particle in momentum space. We obtain

$$\frac{\partial}{\partial t}\mathcal{E}\{\hat{\rho}(\underline{p},\underline{p},t;\alpha)\} = \int_{R^3} d\underline{p}' W(\underline{p},\underline{p}') \times [\mathcal{E}\{\hat{\rho}(\underline{p}',\underline{p}',t;\alpha)\} - \mathcal{E}\{\hat{\rho}(\underline{p},\underline{p},t;\alpha)\}]. \quad (4.10)$$

This equation, with the transition probability W given by (4.7), is identical to the result recently obtained by Papanicolaou.[2] A master equation which is uniformly valid in time may be derived from (4.5) and (4.1). After integrating over $\underline{x}$-space, one has

$$\frac{\partial}{\partial t}\mathcal{E}\{\hat{\rho}(\underline{p},\underline{p},t;\alpha)\} = \frac{2}{\hbar^2}\int_{R^3} d\underline{p}' \int_0^t d\tau Q(\underline{p},\underline{p}',\tau) \times [\mathcal{E}\{\hat{\rho}(\underline{p}',\underline{p}',t-\tau;\alpha)\} - \mathcal{E}\{\hat{\rho}(\underline{p},\underline{p},t-\tau;\alpha)\}]. \quad (4.11)$$

In the long-time Markovian approximation, Eq. (4.11) reduces to Papanicolaou's result, Eq. (4.10).

5. KINETIC EQUATIONS IN SPECIAL CASES

In this section, we shall derive the explicit form of the kinetic equation in the long-time Markovian limit for several special kinds of fluctuations and correlation functions. Analogous kinetic equations which are uniformly valid in time may also be derived.

Case (i): δV has δ-function correlations in time.

Assuming that $\Gamma(\underline{y},\tau) = 2\pi\hbar\gamma(\underline{y})\delta(\tau)$, we have $\hat{\Gamma}(\underline{p},u) = \hat{\gamma}(\underline{p})$, where $\hat{\gamma}(\underline{p})$ is the Fourier transform of $\gamma(\underline{y})$, defined analogously to Eq. (4.3b). The kinetic equation (4.6) then becomes

$$\left(\frac{\partial}{\partial t} + \frac{1}{m}\underline{p}\cdot\frac{\partial}{\partial \underline{x}}\right)\mathcal{E}\{f(\underline{x},\underline{p},t;\alpha)\} = \frac{2\pi}{\hbar}\int_{R^3} d\underline{p}'\hat{\gamma}(\underline{p}-\underline{p}') \times [\mathcal{E}\{f(\underline{x},\underline{p}',t;\alpha) - \mathcal{E}\{f(\underline{x},\underline{p},t;\alpha)\}]. \quad (5.1)$$

The *scattering rate* (also called the *extinction coefficient* or *collision frequency*) is defined in general by

$$\nu(\underline{p}) = \int_{R^3} d\underline{p}' W(\underline{p},\underline{p}'). \quad (5.2)$$

In this case, the scattering rate is independent of $\underline{p}$ and is given by

$$\nu = \frac{2\pi}{\hbar}\int_{R^3} d\underline{p}'\hat{\gamma}(\underline{p}-\underline{p}') = \frac{2\pi}{\hbar}\gamma(0). \quad (5.3)$$

In the Appendix it will be shown by means of the Novikov functional method that for a potential field which constitutes a δ-correlated (in time), homogeneous, wide-sense stationary Gaussian process, the kinetic equation (5.1) for the mean Wigner distribution function is the exact statistical equation.

Case (ii): δV has no time dependence.

Assuming that $\Gamma(\underline{y},\tau) = \gamma(\underline{y})$, we have $\hat{\Gamma}(\underline{p},u) = \hat{\gamma}(\underline{p})\delta(u)$. The transition probability W becomes

$$W(\underline{p},\underline{p}') = \frac{2\pi}{\hbar}\gamma(\underline{p}-\underline{p}')\delta\left(\frac{p^2}{2m} - \frac{p'^2}{2m}\right). \quad (5.4)$$

Because of the δ-function in Eq. (5.4), the motion in the phase space takes place on the energy surfaces $p^2/2m = \text{const}$. It is therefore convenient when writing the kinetic equation to replace the vector $\mathbf{p}$ as the independent variable with the unit vector $\hat{\mathbf{p}}$ and $u = p^2/2m$, where $\hat{\mathbf{p}} = \mathbf{p}/|\mathbf{p}|$. Using the relation $d\mathbf{p} = m(2mu)^{1/2}\, d\hat{\mathbf{p}}\, du$, we find that Eq. (4.6) becomes

$$\left[\frac{\partial}{\partial t} + \left(\frac{2u}{m}\right)^{1/2}\hat{\mathbf{p}}\cdot\frac{\partial}{\partial \mathbf{x}}\right]\mathcal{E}\{f(\mathbf{x},\hat{\mathbf{p}},u,t;\alpha)\} = \int_\Omega d\hat{\mathbf{p}}' w(\hat{\mathbf{p}}-\hat{\mathbf{p}}',u)[\mathcal{E}\{f(\mathbf{x},\hat{\mathbf{p}}',u,t;\alpha)\} - \mathcal{E}\{f(\mathbf{x},\hat{\mathbf{p}},u,t;\alpha)\}], \quad (5.5a)$$

$$w(\hat{\mathbf{p}}-\hat{\mathbf{p}}',u) = \frac{m(2mu)^{1/2}}{\hbar^2(2\pi\hbar)^2}\int_{R^3} d\mathbf{y}\gamma(\mathbf{y}) \times \cos[(2mu)^{1/2}\mathbf{y}\cdot(\hat{\mathbf{p}}-\hat{\mathbf{p}}')/\hbar], \quad (5.5b)$$

where Ω denotes the range of $\hat{\mathbf{p}}'$ over the surface of a unit sphere. This kinetic equation has the form of the classical radiation transport equation, in which the energy u (or frequency $\omega = u/\hbar$) appears only as a parameter.

If the fluctuations are isotropic, so that γ depends only on $y = |\mathbf{y}|$, then w may be further simplified to

$$w(\hat{\mathbf{p}}-\hat{\mathbf{p}}',u) = \frac{m}{2\pi\hbar^3}\int_0^\infty dy y\gamma(y)\left(\frac{\sin[2(y/\lambda\!\!\!^{-})\sin(\theta/2)]}{\sin(\theta/2)}\right), \quad (5.6)$$

where $\lambda\!\!\!^{-} = \hbar/(2mu)^{1/2}$ is the reduced Compton wavelength of the particle with energy u, and θ is defined by $\cos\theta = \hat{\mathbf{p}}\cdot\hat{\mathbf{p}}'$. It should be noted that the assumption of isotropic fluctuations does not imply isotropic scattering. However, when the Compton wavelength is much

greater than the correlation scale lengths, then the scattering does become isotropic since Eq. (5.6) reduces to

$$w(\hat{\mathbf{p}}-\hat{\mathbf{p}}',u)=\frac{m}{\pi\lambda\hbar^3}\int_0^\infty dy y^2\gamma(y). \tag{5.7}$$

The kinetic equation (5.5) then becomes

$$\left[\frac{1}{v(u)}\frac{\partial}{\partial t}+\hat{\mathbf{p}}\cdot\frac{\partial}{\partial\mathbf{x}}\right]\mathcal{E}\{f(\mathbf{x},\hat{\mathbf{p}},u,t;\alpha)\} = \kappa\left[-\mathcal{E}\{f(\mathbf{x},\hat{\mathbf{p}},u,t;\alpha)\}+\frac{1}{4\pi}\int_\Omega d\hat{\mathbf{p}}'\mathcal{E}\{f(\mathbf{x},\hat{\mathbf{p}}',u,t;\alpha)\}\right], \tag{5.8}$$

where $v(u)=(2u/m)^{1/2}$, and the spatial attenuation coefficient κ is given by

$$\kappa=\frac{4m^2}{\hbar^4}\int_0^\infty dy\, y^2\gamma(y). \tag{5.9}$$

Equation (5.8) has the usual form of the radiation transport equation for isotropic scattering, and many results (analytical and numerical) are known regarding solutions of this equation.

Case (iii): δV has δ-function correlations in space.

Assuming that $\Gamma(\underline{y},\tau)=(2\pi\hbar)^3\delta(\underline{y})\gamma(\tau)$, we find that

$$W(\underline{p},\underline{p}')=\frac{2}{\hbar^2}\int_0^\infty d\tau\gamma(\tau)\cos\left[\tau\left(\frac{p^2}{2m}-\frac{p'^2}{2m}\right)/\hbar\right]. \tag{5.10}$$

Thus, the scattering is isotropic, and it is again [as in Case (ii)] convenient to use $\hat{\mathbf{p}}$ and $u=p^2/2m$ as independent variables in the kinetic equation (4.10). In this case, it turns out that the kinetic equation is given by

$$\left[\frac{\partial}{\partial t}+v(u)\hat{\mathbf{p}}\cdot\frac{\partial}{\partial\mathbf{x}}\right]\mathcal{E}\{f(\mathbf{x},\hat{\mathbf{p}},u,t;\alpha)\} = \int_0^\infty du'v(u')\sigma(u-u')\times\left[-\mathcal{E}\{f(\mathbf{x},\hat{\mathbf{p}},u,t;\alpha)\}-\frac{1}{4\pi}\int_\Omega d\hat{\mathbf{p}}'\mathcal{E}\{f(\underline{x},\hat{\underline{p}}',u',t;\alpha)\}\right], \tag{5.11}$$

where $v(u)=(2u/m)^{1/2}$, and

$$\sigma(u)=\frac{8\pi m^2}{\hbar^2}\int_0^\infty d\tau\gamma(\tau)\cos(u\tau/\hbar). \tag{5.12}$$

Case (iv): δV varies slowly in space.

First we note that if the fluctuations in δV are spatially independent, then $W(\mathbf{p},\mathbf{p}')$ is proportional to $\delta(\mathbf{p}-\mathbf{p}')$ and the collision integral [right-hand side of (4.10)] vanishes. We shall, therefore, consider the case where δV varies slowly and space-correlation lengths are assumed to be large compared to Compton wavelengths. This is the correspondence limit, and one expects to obtain a diffusion, or a Fokker–Planck equation.[14] We shall follow the method used by Landau to derive the Fokker–Planck equation for a plasma from the Boltzmann equation.[16]

Denote the right-hand side of (4.10) by $C(\mathbf{p})$:

$$C(\mathbf{p})=\int_{R^3}d\mathbf{p}'W(\mathbf{p},\mathbf{p}')[\mathcal{E}\{f(\underline{p}')\}-\mathcal{E}\{f(\underline{p})\}], \tag{5.13}$$

where the other arguments of f have been suppressed for notational convenience. Let $\mathbf{p}'=\mathbf{p}+\Delta\mathbf{p}$ and expand the factor within the square brackets to get

$$C(\mathbf{p})=\int_{R^3}d\Delta\mathbf{p}W(\mathbf{p},\mathbf{p}+\Delta\mathbf{p})\times\left[\Delta p\cdot\frac{\partial}{\partial\mathbf{p}}\mathcal{E}\{f(\mathbf{p})\}+\frac{1}{2}\Delta\mathbf{p}\Delta\mathbf{p}:\frac{\partial}{\partial\mathbf{p}}\frac{\partial}{\partial\mathbf{p}}\mathcal{E}\{f(\mathbf{p})\}\right]. \tag{5.14}$$

To first order in $\Delta\mathbf{p}$, one has

$$W(\mathbf{p},\mathbf{p}+\Delta\mathbf{p})\simeq W(\mathbf{p}-\tfrac{1}{2}\Delta\mathbf{p},\mathbf{p}+\tfrac{1}{2}\Delta\mathbf{p})+\frac{1}{2}\Delta\mathbf{p}\cdot\frac{\partial}{\partial\mathbf{p}}W(\mathbf{p}-\tfrac{1}{2}\Delta\mathbf{p},\mathbf{p}+\tfrac{1}{2}\Delta\mathbf{p}). \tag{5.15}$$

Introducing (5.15) into (5.14) and rearranging terms gives

$$C(\mathbf{p})=\int_{R^3}d\Delta\mathbf{p}W(\mathbf{p}-\tfrac{1}{2}\Delta\mathbf{p},\mathbf{p}+\tfrac{1}{2}\Delta\mathbf{p})\Delta\mathbf{p}\cdot\frac{\partial}{\partial\mathbf{p}}\mathcal{E}\{f(\mathbf{p})\}+\frac{1}{2}\int_{R^3}d\Delta\mathbf{p}\Delta\mathbf{p}\cdot\frac{\partial}{\partial\mathbf{p}}W(\mathbf{p}-\tfrac{1}{2}\Delta\mathbf{p},\mathbf{p}+\tfrac{1}{2}\Delta\mathbf{p})\Delta\mathbf{p}\cdot\frac{\partial}{\partial\mathbf{p}}\mathcal{E}\{f(\mathbf{p})\}. \tag{5.16}$$

Since $W(\mathbf{p}-\frac{1}{2}\Delta\mathbf{p},\mathbf{p}+\frac{1}{2}\Delta\mathbf{p})$ is an even function of $\Delta\mathbf{p}$ by virtue of the principle of detailed balance, the first term on the right-hand side of (5.16) vanishes. The remaining term represents a diffusion operator in momentum space, and the full kinetic equation becomes

$$\left[\frac{\partial}{\partial t}+\frac{1}{m}\underline{p}\cdot\frac{\partial}{\partial\mathbf{x}}\right]\mathcal{E}\{f(\underline{x},\underline{p},t;\alpha)\}=\frac{\partial}{\partial\underline{p}}\cdot\underline{\underline{D}}(\underline{p})\cdot\frac{\partial}{\partial\underline{p}}\mathcal{E}\{f(\underline{x},\underline{p},t;\alpha)\}, \tag{5.17}$$

with the diffusion dyadic $\underline{\underline{D}}$ given by

$$\underline{\underline{D}}(\underline{p})=\frac{1}{2}\int d\Delta\underline{p}\Delta\underline{p}\Delta\underline{p}W(\underline{p}-\tfrac{1}{2}\Delta\underline{p},\underline{p}+\tfrac{1}{2}\Delta\underline{p}). \tag{5.18}$$

Using Eqs. (4.7a) and (4.3b), this can be written in terms of Γ as follows:

$$\underline{\underline{D}}(\underline{p})=\frac{\pi}{\hbar}\int_{R^3}d\Delta\underline{p}\Delta\underline{p}\Delta\underline{p}\hat{\Gamma}(\Delta\underline{p},\tfrac{1}{m}\underline{p}\cdot\Delta\underline{p}) =-\int_{R^3}d\underline{y}\int_0^\infty d\tau\,\delta\left(\underline{y}-\frac{\tau}{m}\underline{p}\right)\frac{\partial}{\partial\underline{y}}\frac{\partial}{\partial\underline{y}}\Gamma(\underline{y},\tau). \tag{5.19}$$

In Sec. 6, we shall show that the Fokker–Planck equation (5.17) may also be derived by applying the method of smoothing to the stochastic Liouville equation. There the diffusion coefficient will be discussed in more detail.

6. KINETIC THEORY FOR THE STOCHASTIC LIOUVILLE EQUATION

Under the assumptions that the background potential field is homogeneous and stationary, one has, in the framework of the first-order smoothing approximation [cf. Eq. (3.2)], the following kinetic equation associated with the stochastic Liouville equation:

$$\left[\frac{\partial}{\partial t}+\frac{1}{m}\underline{p}\cdot\frac{\partial}{\partial\underline{x}}\right]\mathcal{E}\{f(\underline{x},\underline{p},t;\alpha)\}=\Theta\mathcal{E}\{f(\underline{x},\underline{p},t;\alpha)\}, \tag{6.1a}$$

$$\Theta\mathcal{E}\{f(\underline{x},\underline{p},t;\alpha)\} =\frac{\partial}{\partial\underline{p}}\cdot\left[\int_0^t d\tau\,\mathcal{E}\left\{\frac{\partial}{\partial\underline{x}}\delta V(\underline{x},t;\alpha)\frac{\partial}{\partial\underline{x}}\delta V\left(\underline{x}-\frac{\tau}{m}\underline{p},t-\tau;\alpha\right)\right\}\cdot\frac{\partial}{\partial\underline{p}}\mathcal{E}\left\{f\left(\underline{x}-\frac{\tau}{m}\underline{p},\underline{p},t-\tau;\alpha\right)\right\}\right]. \tag{6.1b}$$

For random fluctuations which are statistically homogeneous, *wide-sense* stationary, and, in addition, δ-correlated in time, one can easily carry out the time integration in (6.1), with the result

$$\Theta\mathcal{E}\{f(\underline{x},\underline{p},t;\alpha)\} = \frac{\partial}{\partial \underline{p}}\cdot\underline{\underline{D}}\cdot\frac{\partial}{\partial \underline{p}}\mathcal{E}\{f(\underline{x},\underline{p},t;\alpha)\}, \tag{6.2a}$$

$$\underline{\underline{D}} = -\tfrac{1}{2}\lim_{y\to 0}\left[\frac{\partial}{\partial \underline{y}}\frac{\partial}{\partial \underline{y}}\gamma(\underline{y})\right]. \tag{6.2b}$$

Here, $\gamma(\underline{y})$ is defined by $\Gamma(\underline{y},\tau) = \gamma(\underline{y})\delta(\tau)$. It is seen, therefore, that in this case (6.1) becomes a Fokker–Planck equation with a dyadic diffusion coefficient which is independent of momentum. In the Appendix, it will be shown by a functional approach, that if $\delta V(\underline{x},t;\alpha)$ is a δ-correlated (in time), homogeneous, wide-sense stationary, Gaussian process, then (6.1a), with $\Theta\mathcal{E}\{f(\underline{x},\underline{p},t;\alpha)\}$ given in (6.2), is the exact statistical equation for the mean Wigner distribution function.

In the long-time, Markovian, diffusion approximation, (6.1) simplifies to

$$\left[\frac{\partial}{\partial t} + \frac{1}{m}\underline{p}\cdot\frac{\partial}{\partial \underline{x}}\right]\mathcal{E}\{f(\underline{x},\underline{p},t;\alpha)\} = \frac{\partial}{\partial \underline{p}}\cdot\left[\underline{\underline{D}}(\underline{p})\cdot\frac{\partial}{\partial \underline{p}}\mathcal{E}\{f(\underline{x},\underline{p},t;\alpha)\}\right]. \tag{6.3}$$

This is a Fokker–Planck equation in phase-space. The dyadic diffusion coefficient $\underline{\underline{D}}$ is given by

$$\underline{\underline{D}}(\underline{p}) = \int_0^\infty d\tau\,\mathcal{E}\left\{\frac{\partial}{\partial \underline{x}}\delta V(\underline{x},t;\alpha)\frac{\partial}{\partial \underline{x}}\delta V\left(\underline{x}-\frac{\tau}{m}\underline{p},t-\tau;\alpha\right)\right\} \tag{6.4a}$$

$$= -\int_{R^3} d\underline{y}\int_0^\infty d\tau\,\delta\left(\underline{y}-\frac{\tau}{m}\underline{p}\right)\frac{\partial}{\partial \underline{y}}\frac{\partial}{\partial \underline{y}}\Gamma(\underline{y},\tau). \tag{6.4b}$$

Equations (6.3) and (6.4b) were previously derived in Sec. 5, Case (iv). For spatially homogeneous and isotropic, and wide-sense stationary random fluctuations, (6.4b) becomes

$$\underline{\underline{D}}(\underline{p}) = \frac{m}{p}\int_0^\infty dy\left[(\underline{\underline{I}}-\hat{\underline{p}}\hat{\underline{p}})\frac{1}{y}\frac{\partial}{\partial y}\Gamma\left(y,\frac{m}{p}y\right) + \hat{\underline{p}}\hat{\underline{p}}\frac{\partial^2}{\partial y^2}\Gamma\left(y,\frac{m}{p}y\right)\right], \tag{6.5}$$

where $\partial/\partial y$ signifies that derivatives are to be taken with respect to only the first argument of Γ and $p = |\underline{p}|$, $\hat{\underline{p}} = \underline{p}/p$. Assuming that

$$\frac{\partial}{\partial y}\Gamma\left(y,\frac{m}{p}y\right) \to 0 \quad \text{as } y\to 0 \text{ and } y\to\infty,$$

we find that (6.5) simplifies to

$$\underline{\underline{D}}(\underline{p}) = \frac{m}{p}\frac{1}{l(p)}(\underline{\underline{I}}-\hat{\underline{p}}\hat{\underline{p}}), \tag{6.6a}$$

$$\frac{1}{l(p)} = -\int_0^\infty dy\frac{1}{y}\frac{\partial}{\partial y}\Gamma\left(y,\frac{m}{p}y\right). \tag{6.6b}$$

The parameter $l(p)$ introduced by means of the integral (6.6b) is related to the correlation length and the correlation time of the random process $\delta V(\underline{x},t;\alpha)$.

Introducing (6.6) into (6.3) one has finally the kinetic equation

$$\left[\frac{\partial}{\partial t} + \frac{1}{m}\underline{p}\cdot\frac{\partial}{\partial \underline{x}}\right]\mathcal{E}\{f(\underline{x},\underline{p},t;\alpha)\} = \frac{m}{p^3}\frac{1}{l(p)}\left(\underline{p}\times\frac{\partial}{\partial \underline{p}}\right)^2\mathcal{E}\{f(\underline{x},\underline{p},t;\alpha)\}. \tag{6.7}$$

The left-hand side describes convection with "group velocity" $\underline{p}/m$, whereas the right-hand side describes angular diffusion in momentum space.

When the Fokker–Planck equation (6.7) is considered in a three-dimensional space, it is often convenient to introduce a spherical polar coordinate system in momentum space: $\underline{p} = (p,\theta,\phi)$. Then, we use (6.7) with

$$\left(\underline{p}\times\frac{\partial}{\partial \underline{p}}\right)^2 = \frac{1}{\sin^2\theta}\frac{\partial^2}{\partial\phi^2} + \frac{1}{\sin\theta}\frac{\partial}{\partial\theta}\left(\sin\theta\frac{\partial}{\partial\theta}\right). \tag{6.8}$$

In closing this section, we wish to point out that in the special case that δV is independent of time, and for the long-time, Markovian, diffusion approximation, the mean Wigner distribution function obeys, again, the kinetic equation (6.7), with the only difference that the parameter l is now independent of momentum and is given explicitly as follows[13]:

$$\frac{1}{l} = -\int_0^\infty dy\frac{1}{y}\frac{\partial}{\partial y}\gamma(y). \tag{6.9}$$

APPENDIX: THE NOVIKOV FUNCTIONAL FORMALISM

We shall establish in the appendix that if the potential field is a δ-correlated (in time), homogeneous, wide-sense stationary, Gaussian process, the kinetic equations (5.1) and (6.2) for the stochastic Wigner and Liouville equations are the exact statistical equations for the mean Wigner distribution function. Towards this goal we shall use a functional approach introduced first by Novikov (cf. Ref. 4) in the study of turbulence theory. More recently the same technique has been used by Tatarskii and Klyatskin[17–20] who studied wave propagation in a random half-space by approximating the scalar Helmholtz equation by a (complex) parabolic equation (a Schrödinger-like equation).

A. The stochastic Wigner equation

We commence with the stochastic Wigner equation (2.9) with $\theta f(\underline{x},\underline{p},t;\alpha)$ given by the representation (2.10b), viz.,

$$\left[\frac{\partial}{\partial t} + \frac{1}{m}\underline{p}\cdot\frac{\partial}{\partial \underline{x}}\right]f(\underline{x},\underline{p},t;\alpha) = \theta f(\underline{x},\underline{p},t;\alpha), \tag{A1a}$$

$$\theta f(\underline{x},\underline{p},t;\alpha) = (i\hbar)^{-1}(2\pi\hbar)^{-3}\int_{R^3} d\underline{y}\,e^{i\underline{p}\cdot\underline{y}/\hbar} \times \rho(\underline{x}+\tfrac{1}{2}\underline{y},\underline{x}-\tfrac{1}{2}\underline{y},t;\alpha) \times [\delta V(\underline{x}-\tfrac{1}{2}\underline{y},t;\alpha) - \delta V(\underline{x}+\tfrac{1}{2}\underline{y},t;\alpha)]. \tag{A1b}$$

Averaging both sides of this equation shows that one has to determine the quantity $\mathcal{E}\{\rho(\underline{x}+\frac{1}{2}\underline{y},\underline{x}-\frac{1}{2}\underline{y},t;\alpha)[\delta V(\underline{x}-\frac{1}{2}\underline{y},t;\alpha) - \delta V(\underline{x}+\frac{1}{2}\underline{y},t;\alpha)]\}$ in terms of the statistical properties of the random process $\delta V(\underline{x},t;\alpha)$ and the ensemble average of the density function. Under the aforementioned properties of the process $\delta V(\underline{x},t;\alpha)$, viz.,

$$\mathcal{E}\{\delta V(\underline{x}_2,t_2;\alpha)\delta V(\underline{x}_1,t_1;\alpha)\} = 2\pi\hbar\delta(t_2-t_1)\gamma(\underline{x}_2-\underline{x}_1), \tag{A2}$$

one has that

$$\mathcal{E}\{\rho(\underline{x}_2,\underline{x}_1,t;\alpha)[\delta V(\underline{x}_1,t;\alpha) - \delta V(\underline{x}_2,t;\alpha)]\}$$

$$= \int_{R^1} dt'\int_{R^3} d\underline{x}'_2\int_{R^3} d\underline{x}'_1\,\mathcal{E}\{[\delta V(\underline{x}_1,t;\alpha) - \delta V(\underline{x}_2,t;\alpha)][\delta V(\underline{x}'_1,t';\alpha) - \delta V(\underline{x}'_2,t';\alpha)]\} \times \mathcal{E}\{\bar{\delta}\rho(\underline{x}_2,\underline{x}_1,t;\alpha)/\bar{\delta}[\delta V(\underline{x}'_1,t';\alpha) - \delta V(\underline{x}'_2,t';\alpha)]\}$$

$$= 2\pi\hbar(i\hbar)^{-1}[\bar{\gamma}(0) - \bar{\gamma}(\underline{x}_2-\underline{x}_1)]\mathcal{E}\{\rho(\underline{x}_2,\underline{x}_1,t;\alpha)\}. \tag{A3}$$

(The symbol $\bar{\delta}(\cdot)$ in (A3) denotes a functional derivative.) This result follows readily from our knowledge of the equation for the density function [cf. Eq. (2.4)]. Using the coordinate transformations $\underline{x}_2 \to \underline{x} + \frac{1}{2}\underline{y}$, $\underline{x}_1 \to \underline{x} - \frac{1}{2}\underline{y}$ in (A3) and introducing the resulting expression into the statistically averaged equation (A1), we obtain

$$\left[\frac{\partial}{\partial t} + \frac{1}{m}\underline{p}\cdot\frac{\partial}{\partial \underline{x}}\right]\mathcal{E}\{f(\underline{x},\underline{p},t;\alpha)\} = -\frac{2\pi}{\hbar}\frac{1}{(2\pi\hbar)^3}\int_{R^3} d\underline{y}\, e^{i\underline{p}\cdot\underline{y}/\hbar}\,[\bar{\gamma}(0) - \bar{\gamma}(\underline{y})] \times \mathcal{E}\{\rho(\underline{x} + \tfrac{1}{2}\underline{y}, \underline{x} - \tfrac{1}{2}\underline{y}, t;\alpha)\}. \quad \text{(A4)}$$

Introducing next the spectrum of the spatial correlation coefficient as previously, and bearing in mind the definition of the Wigner distribution function, (A4) changes to the simple, convolution-type kinetic equation

$$\left[\frac{\partial}{\partial t} + \frac{1}{m}\underline{p}\cdot\frac{\partial}{\partial \underline{x}} - \frac{2\pi}{\hbar}\gamma(0)\right]\mathcal{E}\{f(\underline{x},\underline{p},t;\alpha)\} = \frac{2\pi}{\hbar}\int_{R^3}\hat{\gamma}(\underline{p}-\underline{p}')\mathcal{E}\{f(\underline{x},\underline{p}',t;\alpha)\}. \quad \text{(A5)}$$

It should be noted that

$$\frac{2\pi}{\hbar}\bar{\gamma}(0) = \frac{2\pi}{\hbar}\int_{R^3} d\underline{p}\,\hat{\gamma}(\underline{p}) = \nu. \quad \text{(A6)}$$

It is seen, therefore, that (A5) is identical to the kinetic equation obtained in Sec. 5 [cf. Eq. (5.1)].

B. The stochastic Liouville equation

We begin with the stochastic Liouville equation [cf. Eqs. (2.8), (2.10)] and average both sides:

$$\left[\frac{\partial}{\partial t} + \frac{1}{m}\underline{p}\cdot\frac{\partial}{\partial \underline{x}}\right]\mathcal{E}\{f(\underline{x},\underline{p},t;\alpha)\} = \sum_{j=1}^{3}\frac{\partial}{\partial p_j}\,\mathcal{E}\left\{\left[\frac{\partial}{\partial x_j}\delta V(\underline{x},t;\alpha)f(\underline{x},\underline{p},t;\alpha)\right]\right\}. \quad \text{(A7)}$$

Proceeding along the lines of Novikov, it is found that for the type of random fluctuations considered here, one has

$$\begin{aligned}\mathcal{E}\left\{\frac{\partial}{\partial x_j}\delta V(\underline{x},t;\alpha)f(\underline{x},\underline{p},t;\alpha)\right\} &= \sum_{k=1}^{3}\int_{R^1}dt'\int_{R^3}d\underline{x}'\,\mathcal{E}\left\{\frac{\partial}{\partial x_j}\delta V(\underline{x},t;\alpha)\frac{\partial}{\partial x_k}\delta V(\underline{x}',t';\alpha)\right. \\ &\quad \left.\times\frac{\mathcal{E}\{\bar{\delta}f(\underline{x},\underline{p},t;\alpha)\}}{\bar{\delta}[(\partial/\partial x_k')\delta V(\underline{x}',t';\alpha)]}\right\} \\ &= \sum_{k=1}^{3}\int_{R^3}d\underline{x}'\left[\frac{\partial^2}{\partial x_j\,\partial x_k'}\gamma(\underline{x}-\underline{x}')\right]\mathcal{E}\left\{\frac{\bar{\delta}f(\underline{x},\underline{p},t;\alpha)}{\bar{\delta}[(\partial/\partial x_k')\delta V(\underline{x}',t;\alpha)]}\right\}.\end{aligned} \quad \text{(A8)}$$

To proceed further, we shall have to compute the functional derivative $\bar{\delta}f(\underline{x},\underline{p},t;\alpha)/\bar{\delta}[(\partial/\partial x_k')\delta V(\underline{x}',t;\alpha)]$. Towards this end, the unaveraged Liouville equation is integrated over time to obtain

$$f(\underline{x},\underline{p},t;\alpha) - f_0(\underline{x},\underline{p}) + \frac{1}{m}\underline{p}\cdot\frac{\partial}{\partial \underline{x}}\int_0^t d\tau f(\underline{x},\underline{p},\tau;\alpha) = \sum_{j=1}^{3}\frac{\partial}{\partial p_j}\int_0^t d\tau\frac{\partial}{\partial x_j}\delta V(\underline{x},\tau;\alpha)f(\underline{x},\underline{p},\tau;\alpha). \quad \text{(A9)}$$

[$f_0(\underline{x},\underline{p})$ is the initial ($t = 0$) value of the Wigner distribution function.] We introduce next the auxiliary function

$$\psi(\tau) = \int_{-\infty}^{\tau}\delta(\tau')\,d\tau' = \begin{cases}0, & \tau < 0\\ 1/2, & \tau = 0\\ 1, & \tau > 0.\end{cases} \quad \text{(A10)}$$

Adding the factor $\psi(t-\tau)\delta(\underline{x}-\underline{x}'')$ to the integrand on the right-hand side of (A9) changes that equation into the form

$$\begin{aligned}f(\underline{x},\underline{p},t;\alpha) - f_0(\underline{x},\underline{p}) + \frac{1}{m}\underline{p}\cdot\frac{\partial}{\partial \underline{x}}\int_0^t d\tau f(\underline{x},\underline{p},t;\alpha) &= \sum_{j=1}^{3}\frac{\partial}{\partial p_j}\int_0^\infty d\tau\int_{R^3}d\underline{x}''\psi(t-\tau) \\ &\quad\times\delta(\underline{x}-\underline{x}'')\frac{\partial}{\partial x_j''}\delta V(\underline{x}'',\tau;\alpha)f(\underline{x}'',\underline{p},\tau;\alpha).\end{aligned} \quad \text{(A11)}$$

We operate next on both sides of this expression with $\bar{\delta}/\bar{\delta}[(\partial/\partial x_k')\delta V(\underline{x}',t';\alpha)]$ taking into consideration the fact that $\bar{\delta}[(\partial/\partial x_j'')\delta V(\underline{x}'',\tau;\alpha)]/\bar{\delta}[(\partial/\partial x_k')\delta V(\underline{x}',t';\alpha)] = \delta_{jk}\delta(\underline{x}''-\underline{x}')\delta(\tau-t')$:

$$\begin{aligned}\frac{\bar{\delta}f(\underline{x},\underline{p},t;\alpha)}{\bar{\delta}[(\partial/\partial x_k')\delta V(\underline{x}',t';\alpha)]} + \frac{1}{m}\underline{p}\cdot\frac{\partial}{\partial \underline{x}}\int_0^t d\tau\frac{\bar{\delta}f(\underline{x},\underline{p},\tau;\alpha)}{\bar{\delta}[(\partial/\partial x_k')\delta V(\underline{x}',t';\alpha)]} &= (\partial/\partial p_k)[\psi(t-t')\delta(\underline{x}-\underline{x}')f(\underline{x}',\underline{p},t';\alpha)]\cdots \\ &\quad+\sum_{j=1}^{3}\frac{\partial}{\partial p_j}\int_0^t d\tau\int_{R^3}d\underline{x}''\psi(t-\tau) \\ &\quad\times\delta(\underline{x}-\underline{x}'')\frac{\partial}{\partial x_j''}\delta V(\underline{x}'',\tau;\alpha)\frac{\bar{\delta}f(\underline{x}'',\underline{p},\tau;\alpha)}{\bar{\delta}[(\partial/\partial x_k')\delta V(\underline{x}',t';\alpha)]}.\end{aligned} \quad \text{(A12)}$$

It is noted next that changing the lower limit of integration over τ to t' does not alter (A12). We have, then,

$$\frac{\bar{\delta}f(\underline{x},\underline{p},t;\alpha)}{\bar{\delta}[(\partial/\partial x_k')\delta V(\underline{x}',t;\alpha)]} = -\frac{1}{2}\delta(\underline{x}-\underline{x}')\frac{\partial}{\partial p_k}f(\underline{x}',\underline{p},t;\alpha). \quad \text{(A13)}$$

Introducing (A13) into (A8) and performing the integration over x' finally yields the Fokker–Planck equation

$$\left[\frac{\partial}{\partial t} + \frac{1}{m}\underline{p}\cdot\frac{\partial}{\partial \underline{x}}\right]\mathcal{E}\{f(\underline{x},\underline{p},t;\alpha)\} = \frac{\partial}{\partial \underline{p}}\cdot\underline{\underline{D}}\cdot\frac{\partial}{\partial \underline{p}}\,\mathcal{E}\{f(\underline{x},\underline{p},t;\alpha)\}, \quad \text{(A14a)}$$

$$\underline{\underline{D}} = -\tfrac{1}{2}\lim_{\underline{x}\to 0}\left[\frac{\partial}{\partial \underline{x}}\frac{\partial}{\partial \underline{x}}\gamma(\underline{x})\right] \quad \text{(A14b)}$$

which is identical to the kinetic equation (6.2).

ACKNOWLEDGMENTS

The authors wish to express their gratitude to Professor J. B. Keller for his encouragement and interest in this research. They also wish to thank Professors P. L. Chow and G. C. Papanicolaou for several stimulating discussions.

*Research supported in part by the Air Force Office of Scientific Research under Grant No. AFOSR-71-2107.

†Research conducted while the first author was with the Courant Institute and Department of Mathematics, New York University.

[1] S. Chandrasekhar, Rev. Mod. Phys. **15**, 1 (1943); E. Fermi, Astrophys. J. **119**, 1 (1954); R. Kubo, J. Math. Phys. **4**, 174 (1963); P. A. Sturrock, Phys. Rev. **141**, 186 (1966); D. E. Hall and P. A. Sturrock,

Phys. Fluids **10,** 2620 (1967); S. Orszag and R. Kraichnan, Phys. Fluids **10,** 1720 (1967); R. Hakim, J. Math. Phys. **9,** 1805 (1968); A. N. Kaufman, Phys. Fluids **11,** 326 (1968); E. C. Roelof, Can. J. Phys. **46,** S990 (1969); C. E. Newman, P. A. Sturrock, and E. Tademaru, Astrophys. Space Sci. **10,** 102 (1971); G. Papanicolaou, J. Math. Phys. **12,** 1493 (1971); J. R. Jokipii, Astrophys. J. **172,** 319 (1972).

[2]R. J. Glauber, Phys. Rev. **131,** 2766 (1963); B. Crosignani, P. DiPorto, and S. Solimeno, Phys. Rev. **186,** 1342 (1969); B. R. Mollow, Phys. Rev. A **2,** 1477 (1970); G. S. Agarwal, Phys. Rev. A **4,** 739 (1971).

[3]G. Papanicolaou, "A kinetic theory for power transfer in stochastic systems," submitted to J. Math. Phys. (communicated privately to the authors).

[4]E. A. Novikov, Zh. Eksp. Teor. Fiz. **47,** 1919 (1964) [Sov. Phys.-JETP **20,** 1290 (1965)].

[5]E. Wigner, Phys. Rev. **40,** 749 (1932).

[6]H. J. Groenewold, Physica (Utr.) **12,** 405 (1946).

[7]J. E. Moyal, Proc. Camb. Philos. Soc. **45,** 99 (1949).

[8]J. H. Irving and R. W. Zwanzig, J. Chem. Phys. **19,** 1173 (1951).

[9]B. Leaf, J. Math. Phys. **9,** 65 (1968), J. Math. Phys. **9,** 769 (1968).

[10]J. B. Keller, Proc. Symp. Appl. Math. **13,** 277 (1962); Proc. Symp. Appl. Math. **16,** 145 (1964).

[11]V. Frisch, "Wave Propagation in Randon Media" in *Probabilistic Methods in Applied Mathematics*, Vol. 1, edited by A. T. Bharucha-Reid (Academic, New York, 1968).

[12]For a derivation of the first-order smoothing approximation along the lines of the resolvent theory used in nonequilibrium statistical mechanics, see I. M. Besieris, J. Math. Phys. **13,** 358 (1972).

[13]F. D. Tappert and I. M. Besieris, "Stochastic Wave Kinetic Equation and Its Application to Wave Packet Spreading," Proc. Intern. Symp. on Electromagnetic Wave Theory, Tbilisi, U.S.S.R., Sept. 1971.

[14]F. D. Tappert and I. M. Besieris, "Stochastic Wave Kinetic Equations: General Theory," to appear.

[15]See, for example, E. Merzbacher, *Quantum Mechanics* (Wiley, New York, 1961), pp. 466–71.

[16]L. D. Landáu, Zh. Eksp. Teor. Fiz. **7,** 203 (1937) (in Russian); Phys. Z. Sowjetunion **10,** 154 (1936) (in German); *Collected Papers of L. D. Landau,* edited by D. Ter Haar (Gordon and Breach, New York, 1965), pp. 163–70 (in English).

[17] V. I. Tatarskii, Zh. Eksp. Teor. Fiz. **56,** 2106 (1969) [Sov. Phys.-JETP **29,** 1133 (1969)].

[18]V. I. Klyatskin and V. I. Tatarskii, Izv. Vyssh. Uchebn. Zaved. Radiofiz. **13,** 1061 (1970).

[19]V. I. Klyatskin and V. I. Tatarskii, Sov. Phys.-JETP **31,** 335 (1970).

[20]V. I. Klyatskin, Zh. Eksp. Teor. Fiz. **57,** 952 (1969) [Sov. Phys.-JETP **30,** 520 (1970).

Two-magnon bound states in Heisenberg ferromagnets*

A. W. Sáenz and W. W. Zachary

Naval Research Laboratory, Washington, D. C. 20375
(Received 22 June 1972)

A generalized theory of bound two-magnon states in three-dimensional isotropic Heisenberg ferromagnets is given and the passage to the limit in which the total number of spins tends to infinity is handled rigorously. Powerful methods, mostly of the trace-inequality type, are developed for determining upper and lower bounds to the number of such bound states in the latter limit. These methods constitute the central contribution of this paper. In the latter we apply them to investigate the existence of bound two-magnon states in body-centered Heisenberg ferromagnets whose nonvanishing exchange interactions are those of the nearest-neighbor type. In work reported elsewhere, we have employed these methods to study spin-wave impurity states in Heisenberg ferromagnets. They should be useful for determining bounds on the number of localized states in solids in many cases when interactions extending over several orders of neighbors are operative.

1. INTRODUCTION

The main purpose of this paper is to develop a new theoretical approach to the study of localized states in solids, which we shall present here in a form suitable to the study of two-magnon bound states in three-dimensional Heisenberg ferromagnets with isotropic exchange couplings. This isotropy has been assumed purely for the sake of simplicity; it is not essential to the success of the approach. The latter has also been employed to investigate spin-wave impurity states in such ferromagnets,[1] and should have an even wider range of applicability.

Bound magnon states and resonances are of interest for several reasons. In the first place, as is well known, the validity of many theoretical investigations of low-temperature properties of three-dimensional Heisenberg ferromagnets depends critically on the assumption that no low-lying bound states or resonances of n magnons exist ($n = 2, 3, \cdots$).[2,3] Secondly, ferromagnetic bound magnon states are of increasing experimental interest. Silbergliti and Harris[4] have proposed an indirect method for proving their existence experimentally for suitable ferromagnets, namely, by detecting the corresponding resonances. Torrance and Tinkham[5] have actually detected such states in linear chains directly and Thorpe[6] has suggested a direct method for their detection in appropriate cases. Thirdly, currently available theoretical methods for evaluating the energy and momentum parameters of two-magnon resonances are particularly tedious and opaque in the absence of adequate knowledge concerning the corresponding two-magnon bound states, which the methods of this paper are capable of providing in many instances. This knowledge would give insight into the energy and momentum parameters of the resonances in question, and this is an added reason why the study of the above bound states is of interest.

Dyson[7] was the first to publish detailed theoretical considerations on two-magnon bound states in three-dimensional Heisenberg ferromagnets. He concluded that no low-lying bound states exist for cubic ferromagnets of this type whose magnetic ions are coupled solely by isotropic nearest-neighbor (NN) exchange interactions. As is well known, the independent calculation of Hanus[8] and Wortis[9] first established theoretically the existence of two-magnon bound states in this three-dimensional case. For the example of isotropic sc Heisenberg ferromagnets with only NN couplings, their calculations show that, when the total wave-number vector Γ of the bound magnon pair points along the [111] direction,[10] these states occur at energies below the two-magnon continuum, but only for values of Γ fairly close to the zone boundary. For this same example, Wortis[9] also investigated the situation in which the projection of Γ of one of the crystal axes lies on the zone boundary. Moreover, in this last reference he proved that low-lying bound states exist in suitable three-dimensional anisotropic Heisenberg ferromagnets. The above conclusions relative to the [111] direction in the case of the indicated sc ferromagnets are in accord with results reached later by Boyd and Callaway,[11] who considered two-magnon bound states and resonances, and by Shaw.[12] For fcc Heisenberg ferromagnets, isotropically coupled in the manner last stated, Thorpe[6] has shown that such bound states with $\Gamma = 0$ exist above the two-magnon continuum, contrary to what occurs for sc and bcc ferromagnets of this type and mode of coupling.[13]

We proceed to summarize the results of the present paper. Our subsequent statements should be understood in the context of three-dimensional isotropic Heisenberg ferromagnets.

In Sec. 2 we construct a generalized theory of two-magnon bound states, discuss relevant spectral problems in a precise way, and state rigorous conclusions about these states in the limit $N \to \infty$ (Theorem 2.1). This theory applies to spin lattices of the Bravais type, coupled by exchange interactions connecting any finite number of orders of nearest neighbors. In the theory in question, and throughout this paper, a basic role is played by the properties of the pertinent Lippmann–Schwinger two-body kernel, just as in Boyd and Callaway's theory of such bound states,[11] which is generalized by the present investigation.

Sec. 3 is devoted to the development of procedures (Theorems 3.1-3.5), mainly of the trace-inequality type, for determining rigorous upper and lower bounds to the number of two-magnon bound states in this limit. These procedures are based on monotonicity and continuity properties of the eigenvalues of the limiting form of the two-body kernel for $N \to \infty$. Our trace-inequality techniques are similar in spirit to methods which have been stated within the context of bound states of two nonrelativistic particles.

If used in conjunction with modern electronic computers, we believe that the methods of Sec. 3 are well suited to give good estimates of the regions of existence and nonexistence of these bound states, even when distant-range interactions are present.

In Secs. 4 and 5, the methods of Sec. 3 are applied to investigate the occurrence of bound states in bcc ferro-

magnets with purely NN interactions, for vectors Γ of the pertinent Brillouin zone parallel to the [100], [110], and [111] directions. In these calculations, it was only necessary to use the earlier and less efficient exact eigenvalue methods for the study of magnon bound states in very limited ranges of the relevant energy and momentum parameters. All the bound states found in this investigation of bcc ferromagnets occurred for values of Γ in these directions lying fairly close to the zone boundary. While these results give one confidence that no low-lying bound two-magnon states exist for the ferromagnets studied in Secs. 4 and 5, we are certainly not claiming to have proved this. In the latter two sections, trace-inequality estimates of bound state thresholds are compared with values of those thresholds obtained by direct eigenvalue calculations for the [100] and [110] directions of Γ. Good agreement was obtained between these estimates and the results of these eigenvalue calculations for relatively low powers of the matrix entering into the trace-inequality computations. A curious cusp phenomenon for these bcc ferromagnets, which took place for a value of Γ along the [111] direction, and the possible explanation of this phenomenon in terms of a suitably located two-magnon resonance, is mentioned in Sec. 5 and deserves further thought.

During the course of our calculations on the above NN-coupled bcc ferromagnets, we discovered that the degeneracies of the corresponding bound states of magnon pairs could not be accounted for by point group arguments, in sharp contrast with the case of sc ferromagnets with only NN coupling. The underlying group-theoretical explanation of this interesting fact is not known.

The numerical methods used to evaluate certain Green's functions needed for the purposes of Sec. 5 are summarized in the Appendix.

2. GENERALIZED THEORY OF BOUND STATES OF TWO MAGNONS IN HEISENBERG FERROMAGNETS

Let us consider a crystal with N^3 magnetic ions, each having a total spin S. The magnetic interactions among these ions will be assumed to be described by the Hamiltonian

$$H^{(N)}(S) = \frac{1}{2} \sum_{k \neq j} J_{jk}(\mathbf{S}_j \cdot \mathbf{S}_k) + C \sum_j S_j^z. \tag{2.1}$$

Here j and k are the position vectors of the jth and the kth sites of the spin lattice, to which correspond the respective spin vector operators $\mathbf{S}_j$ and $\mathbf{S}_k$, and the summations in (2.1) run over all the magnetic ions. The crystals of interest are assumed to be of Bravais type, to have the familiar parallelipipedal form, and to be subjected to periodic boundary conditions. The constant C is positive and the constants $J_{jk} = J(|\mathbf{j} - \mathbf{k}|)$ are nonnegative. The exchange couplings described by the J_{jk} will be allowed to be nonvanishing only over a finite order of neighbors.

The equations and other statements below involving S are intended to apply to each $S \geq \frac{1}{2}$, unless a statement to the contrary is made. In the considerations of this and of the next section, C, the J_{jk}, and the crystallographic vectors mentioned below will be regarded as fixed, but arbitrary within the limitations stated in the present section.

Boyd and Callaway[11] have introduced a formalism which is particularly appropriate for our investigation of bound states of two magnons. Unfortunately, some of their basic definitions lack precision and this introduces difficulties in deriving certain of their basic conclusions, which we shall require in a generalized form. We proceed to eliminate this deficiency.

For mathematical simplicity, it is assumed that N is an odd positive integer. We denote by L_N the set of all position vectors of the magnetic ions of the crystal, a set which we shall take to consist of all the vectors

$$\sum_{i=1}^{3} m_i \mathbf{a}_i, \qquad |m_i| \leq \tfrac{1}{2}(N-1),$$

which are generated by a basis $\{\mathbf{a}_i\}$. The corresponding infinite lattice, i.e., the set of vectors of this type for which the values of the integers m_i are unrestricted, will be called L.

Let P_N be the set of all vectors

$$2\pi \sum_{i=1}^{3} \frac{n_i}{N} \mathbf{b}_i, \qquad |n_i| \leq \tfrac{1}{2}(N-1),$$

where the n_i are integers and $\{\mathbf{b}_i\}$ is the reciprocal lattice basis corresponding to $\{\mathbf{a}_i\}$. Define Z_N as the set of vectors of the form $\boldsymbol{\lambda} + 2\pi\boldsymbol{\lambda}(\boldsymbol{\gamma})$, where $\boldsymbol{\gamma} \in P_N$ and where at every such $\boldsymbol{\gamma}$ the vector $\boldsymbol{\lambda}(\boldsymbol{\gamma})$ is uniquely determined by the following two requirements. At each $\boldsymbol{\gamma} \in P_N$, $\boldsymbol{\lambda}(\boldsymbol{\gamma})$ is that vector of the reciprocal lattice generated by the $\mathbf{b}_i$ such that $\boldsymbol{\gamma} + 2\pi\boldsymbol{\lambda}(\boldsymbol{\gamma})$ lies in the first Brillouin zone Z pertaining to this lattice. This zone is understood here to be a closed polygonal region of R^3. This choice will prove convenient in our treatment of the limiting case of $N \to \infty$. For each $\boldsymbol{\Gamma} \in Z_N$, we define $Z_N(\boldsymbol{\Gamma})$ as the set of vectors of the form $\boldsymbol{\gamma}_1 + \boldsymbol{\rho}(\boldsymbol{\Gamma})$, where $\boldsymbol{\gamma}_1 \in P_N$ and

$$\boldsymbol{\rho}(\boldsymbol{\Gamma}) \equiv \frac{\pi}{2N} \sum_{i=1}^{3} \left[1 - \cos\left(\frac{N}{2} \boldsymbol{\Gamma} \cdot \mathbf{a}_i\right)\right] \mathbf{b}_i .$$

Naturally, the $\boldsymbol{\gamma}_1$ could have been required to range over Z_N, for example, rather than over P_N, but no particular advantage would be gained thereby in this paper.[14]

Denoting by $|0; S\rangle^{(N)}$ the state of the above ferromagnetic system of N spins in which the latter all point along a given direction, which we shall identify as the negative z direction, and setting $S_{\mathbf{R}}^{+} \equiv S_{\mathbf{R},x} + iS_{\mathbf{R},y}$ for $\mathbf{R} \in L_N$, as usual, we define the two-spin deviation states in terms of this ground state $|0; S\rangle^{(N)}$:

$$\begin{aligned} &|\mathbf{R}_1, \mathbf{R}_2; S\rangle^{(N)} \equiv S_{\mathbf{R}_1}^{+} S_{\mathbf{R}_2}^{+} |0; S\rangle^{(N)}, \\ &|\mathbf{R}_1 + \boldsymbol{\Lambda}_1, \mathbf{R}_2 + \boldsymbol{\Lambda}_2; S\rangle^{(N)} = |\mathbf{R}_1, \mathbf{R}_2; S\rangle^{(N)}, \end{aligned} \tag{2.2}$$

when $\mathbf{R}_1, \mathbf{R}_2 \in L_N$. Here the $\boldsymbol{\Lambda}_i$ are the vectors of the form

$$\boldsymbol{\Lambda} \equiv N \sum_{i=1}^{3} \nu_i \mathbf{a}_i, \tag{2.3}$$

the ν_i being integers.

For each $\mathbf{R} = \sum_{i=1}^{3} m_i \mathbf{a}_i \in L_N$, let $L_N(\mathbf{R})$ be the set of all vectors of the form $\mathbf{R}' + \boldsymbol{\sigma}(\mathbf{R})$, where $\mathbf{R}' \in L_N$ and

$$\boldsymbol{\sigma}(\mathbf{R}) \equiv \sum_{i=1}^{3} r(m_i) \mathbf{a}_i .$$

Given any integer m, $r(m) \equiv 0$ [$r(m) \equiv \frac{1}{2}$] when m is even (odd). For $S > \frac{1}{2}$, we now introduce for every $\boldsymbol{\Gamma} \in Z_N$ and $\mathbf{R} \in L_N$ the ket

$$|\mathbf{\Gamma},\mathbf{R};S\rangle^{(N)} \equiv N^{-\frac{1}{2}} \sum_{\mathbf{C}\in L_N(\mathbf{R})} e^{i\mathbf{\Gamma}\cdot\mathbf{C}}\Delta^{-1}(\mathbf{R})|\mathbf{C}+\tfrac{1}{2}\mathbf{R},\mathbf{C}-\tfrac{1}{2}\mathbf{R};S\rangle^{(N)}$$
$$= |\mathbf{\Gamma},-\mathbf{R};S\rangle^{(N)},$$
where
$$\Delta(\mathbf{R};S) \equiv [2S(2S-\delta_{\mathbf{R},\mathbf{0}})(1+\delta_{\mathbf{R},\mathbf{0}})]^{1/2},\ \mathbf{R}\in L_N. \tag{2.4}$$

The first definition in (2.4) also applies when $S=\frac{1}{2}$ in the case of each $\mathbf{\Gamma}\in Z_N$, provided that $\mathbf{R}$ iz a nonzero vector of L_N, this definition clearly making no sense for this spin value if $\mathbf{R}=0$, since $\Delta(0;\frac{1}{2})=0$. This difficulty reflects the familiar fact that two-spin-deviation states with two reversed spins at a given lattice site vanish when $S=\frac{1}{2}$.

The kets $|\mathbf{\Gamma},\mathbf{R};S\rangle^{(N)}$ just defined for $S\geq\frac{1}{2}$ will be called "physical" mixed kets. To treat the cases $S=\frac{1}{2}$ and $S>\frac{1}{2}$ on an equal footing, we adjoin to these kets in the case $S=\frac{1}{2}$ a "nonphysical" mixed ket $|\mathbf{\Gamma},0;\frac{1}{2}\rangle^{(N)}$ for each $\mathbf{\Gamma}\in Z_N$. We shall *require* that the orthonormality relations

$$^{(N)}\langle\mathbf{\Gamma}',\mathbf{R}';S\,|\,\mathbf{\Gamma},\mathbf{R};S\rangle^{(N)} = \frac{\delta_{\mathbf{\Gamma}',\mathbf{\Gamma}}(\delta_{\mathbf{R}',\mathbf{R}}+\delta_{\mathbf{R}',-\mathbf{R}})}{(1+\delta_{\mathbf{R},\mathbf{0}})}, \tag{2.5}$$

which are consequences of (2.2) and (2.4) for the physical mixed kets pertaining to each $S\geq\frac{1}{2}$ for $\mathbf{\Gamma},\mathbf{\Gamma}'\in Z_N$ if $\mathbf{R},\mathbf{R}'\in L_N$, also hold when one or both of the mixed kets involved are nonphysical. Hence (2.5) holds for $S\geq\frac{1}{2}$ whenever $\mathbf{\Gamma},\mathbf{\Gamma}',\mathbf{R}$, and $\mathbf{R}'$ fulfill each of these respective requirements.

Let us extend the definitions (2.4) from L_N to L; that of $\Delta(\mathbf{R};\mathbf{S})$ by setting $\Delta(\mathbf{R}+\mathbf{\Lambda};S)\equiv\Delta(\mathbf{R};S)$ for each $\mathbf{R}\in L_N$, where $\mathbf{\Lambda}$ is a vector of the type (2.3), and that of $|\mathbf{\Gamma},\mathbf{R};S\rangle^{(N)}$ by writing

$$|\mathbf{\Gamma},\mathbf{R}+\mathbf{\Lambda};S\rangle^{(N)} \equiv e^{-(1/2)i(\mathbf{\Gamma}\cdot\mathbf{\Lambda})}|\mathbf{\Gamma},\mathbf{R};S\rangle^{(N)}, \tag{2.6}$$

the latter definition applying when $\mathbf{R}$ and $\mathbf{\Lambda}$ are as we have just specified. From these extended definitions, together with (2.2) and (2.4), we infer that

$$|\mathbf{R}_1,\mathbf{R}_2;S\rangle^{(N)} = N^{-1/2}\sum_{\mathbf{\Gamma}\in Z_N} e^{-(1/2)i\mathbf{\Gamma}\cdot(\mathbf{R}_1-\mathbf{R}_2)} \times \Delta(\mathbf{R}_1-\mathbf{R}_2;S)\,|\mathbf{\Gamma},\mathbf{R}_1-\mathbf{R}_2;S\rangle^{(N)} \tag{2.7}$$

for $S\geq\frac{1}{2}$ if $\mathbf{R}_1,\mathbf{R}_2\in L$. The fact that (2.7) holds for this range of $\mathbf{R}_1$ and $\mathbf{R}_2$, and not merely when $\mathbf{R}_1,\mathbf{R}_2\in L_N$, is important in our derivation of the generalized Boyd-Callaway formalism.

It is also useful to define the kets

$$|\mathbf{\Gamma},\boldsymbol{\tau};S\rangle^{(N)} \equiv N^{-1/2}\sum_{\mathbf{R}\in L_N}[1+\delta_{\mathbf{R},\mathbf{0}}]^{1/2}\,e^{i\boldsymbol{\tau}\cdot\mathbf{R}}|\mathbf{\Gamma},\mathbf{R};S\rangle^{(N)} \tag{2.8}$$

at these pairs $(\mathbf{\Gamma},\boldsymbol{\tau})$ at which $\mathbf{\Gamma}\in Z_N$ and $\boldsymbol{\tau}\in Z_N(\mathbf{\Gamma})$. As is the case for two kets of the type $|\mathbf{\Gamma},\mathbf{R};S\rangle^{(N)}$ $(\mathbf{\Gamma}\in Z_N,\mathbf{R}\in L_N)$ labeled with the same S, two kets $|\mathbf{\Gamma},\boldsymbol{\tau};S\rangle^{(N)}[\mathbf{\Gamma}\in Z_N,\boldsymbol{\tau}\in Z_N,(\mathbf{\Gamma})]$ with the same S are either identical or orthogonal. We mention in passing that the property $^{(N)}\langle\mathbf{\Gamma},\boldsymbol{\tau};S\,|\,\mathbf{\Gamma},\boldsymbol{\tau};S\rangle^{(N)}=1$ does not generally hold over the set of values on which the kets (2.8) were defined, but this is immaterial for our purposes.

Let $\mathcal{H}^{(N)}(S)$ be the Hilbert space spanned by the set $\{|\mathbf{R}_1,\mathbf{R}_2;S\rangle^{(N)}:\mathbf{R}_1,\mathbf{R}_2\in L_N\}$ of two-spin-deviation kets, this space also being spanned by all the physical mixed kets with the value of S in question, and let $\hat{\mathcal{H}}^{(N)}(S)$ be the Hilbert space spanned by the set $\{|\mathbf{\Gamma},\mathbf{R};S\rangle^{(N)}:\mathbf{\Gamma}\in Z_N,\mathbf{R}\in L_N\}$. It can be shown that $\hat{\mathcal{H}}^{(N)}(S)$ is also spanned by the set $\{|\mathbf{\Gamma},\boldsymbol{\tau};S\rangle^{(N)}:\mathbf{\Gamma}\in Z_N,\boldsymbol{\tau}\in Z_N(\mathbf{\Gamma})\}$. It is evident that $\hat{\mathcal{H}}^{(N)}(S)=\mathcal{H}^{(N)}(S)$ when $S>\frac{1}{2}$ and that $\hat{\mathcal{H}}^{(N)}(\frac{1}{2})$ is the direct sum of $\mathcal{H}^{(N)}(\frac{1}{2})$ and the subspace spanned by all the unphysical mixed kets with $S=\frac{1}{2}$.

Because of (2.4) and (2.5), it is clear that the set $\{|\mathbf{\Gamma},\mathbf{R};S\rangle^{(N)}:\mathbf{\Gamma}\in Z_N,\mathbf{R}\in L'_N\}$ constitutes an orthogonal basis for $\hat{\mathcal{H}}^{(N)}(S)$. Here L'_N is a subset of L_N containing $\mathbf{R}=0$ and exactly one of each pair $\mathbf{R},-\mathbf{R}$ of nonzero vectors of L_N.

It can now be rigorously shown, employing in particular (2.1), (2.2), (2.4), (2.7), (2.8) and its inverse, and the operator techniques of Ref. 11 that

$$H^{(N)}(S)\,|\mathbf{\Gamma},\mathbf{R};S\rangle^{(N)} = -\sum_{\mathbf{R}'\in L_N} J_{\mathbf{R}'\mathbf{0}}\,\frac{(2S-\delta_{\mathbf{R},\mathbf{0}})\Delta(\mathbf{R};S)}{\Delta(\mathbf{R}+\mathbf{R}';S)} \times\cos(\tfrac{1}{2}\mathbf{\Gamma}\cdot\mathbf{R}')\,|\mathbf{\Gamma},\mathbf{R}+\mathbf{R}';S\rangle^{(N)} - J_{\mathbf{R}\mathbf{0}}|\mathbf{\Gamma},\mathbf{R};S\rangle^{(N)} \tag{2.9}$$

whenever $|\mathbf{\Gamma},\mathbf{R};S\rangle^{(N)}\in\mathcal{H}^{(N)}(S)$, where here and henceforth we simplify matters by redefining $H^{(N)}(S)$ by adding to this operator, as previously defined, a term representing a constant energy shift.

Again to unify the treatment of the cases $S=\frac{1}{2}$ and $S>\frac{1}{2}$, we define as follows at each $S\geq\frac{1}{2}$ an auxiliary operator $h^{(N)}(S)$, whose spectrum differs trivially from that of $H^{(N)}(S)$, when the latter operator is restricted to $\mathcal{H}^{(N)}(S)$. At each spin value, $h^{(N)}(S)$ is that self-adjoint operator with domain $\hat{\mathcal{H}}^{(N)}(S)$ which yields the result shown on the right-hand side of (2.9), when acting on each *physical* or *unphysical* mixed ket with the value of S in question. From the form of this right-hand side, it then follows that at every S value $h^{(N)}(S)$ annihilates every *unphysical* mixed ket labeled with the same S. For $S\geq\frac{1}{2}$, we thus have on $\mathcal{H}^{(N)}(S)$ and its orthogonal complement $\mathcal{H}^{(N)}_{\perp}(S)$ with respect to $\hat{\mathcal{H}}^{(N)}(S)$

$$h^{(N)}(S)\varphi = H^{(N)}(S)\varphi,\quad \varphi\in\mathcal{H}^{(N)}(S),$$
$$h^{(N)}(S)\varphi = 0,\quad \varphi\in\mathcal{H}^{(N)}_{\perp}(S). \tag{2.10}$$

We express $h^{(N)}(S)$ as follows:

$$h^{(N)}(S) = h_0^{(N)}(S) + v^{(N)}(S), \tag{2.11}$$

where $h_0^{(N)}(S)$ and $v^{(N)}(S)$ are self-adjoint operators with domain $\hat{\mathcal{H}}^{(N)}(S)$. In particular,

$$h_0^{(N)}\,|\mathbf{\Gamma},\mathbf{R};\mathbf{S}\rangle^{(N)} = -2S\sum_{\mathbf{R}'\in L_N} J_{\mathbf{R}'\mathbf{0}} \times\cos(\tfrac{1}{2}\mathbf{\Gamma}\cdot\mathbf{R})\left[\frac{1+\delta_{\mathbf{R},-\mathbf{R}'}}{1+\delta_{\mathbf{R},\mathbf{0}}}\right]^{1/2}|\mathbf{\Gamma},\mathbf{R}+\mathbf{R}';S\rangle^{(N)} \tag{2.12}$$

for any $|\mathbf{\Gamma},\mathbf{R};S\rangle^{(N)}\in\hat{\mathcal{H}}^{(N)}(S)$. The operator $h_0^{(N)}(S)$ describes two free magnons and $v^{(N)}(S)$, to which we shall return below, describes their interactions within the present model.

Henceforth we shall not indicate the dependence on S explicitly in most cases, writing, for example, $H^{(N)}$, $|\mathbf{\Gamma},\mathbf{R}\rangle^{(N)}$, and $|\mathbf{\Gamma},\boldsymbol{\tau}\rangle^{(N)}$ instead of $H^{(N)}(S)$, $|\mathbf{\Gamma},\mathbf{R};S\rangle^{(N)}$, and $|\mathbf{\Gamma},\boldsymbol{\tau};S\rangle^{(N)}$, respectively. In the remainder of this section and in the following two sections, we shall suppose that S has a fixed value $S\geq\frac{1}{2}$, and therefore this omission will cause no ambiguities in these three sections. Adequate logical safeguards have been set up to prevent any ambiguities from arising elsewhere in this paper because of this omission.

Notice that $e^{i\tau\cdot\mathbf{R}}|\mathbf{\Gamma},\mathbf{R}\rangle^{(N)}[\mathbf{\Gamma}\in Z_N, \tau\in Z_N(\mathbf{\Gamma}), \mathbf{R}\in L]$ has the periodicity of L, as can be proved by using, in particular, (2.6). Combining this periodicity property with (2.8) and its inverse and with (2.12), we conclude that

$$h_0^{(N)}|\mathbf{\Gamma},\tau\rangle^{(N)} = \epsilon^{(N)}(\mathbf{\Gamma},\tau)|\mathbf{\Gamma},\tau\rangle^{(N)} \tag{2.13}$$

for each $|\mathbf{\Gamma},\tau\rangle^{(N)}\in\hat{\mathcal{K}}^{(N)}$, where

$$\epsilon^{(N)}(\mathbf{\Gamma},\tau) \equiv -4S\sum_{l=1}^{n} J_l\cos(\tfrac{1}{2}\mathbf{\Gamma}\cdot\mathbf{R})\cos(\tau\cdot\mathbf{R}_l). \tag{2.14}$$

The vectors $\mathbf{R}_l, -\mathbf{R}_l (l=1,\ldots,n)$ are the nonzero position vectors of the $2n$ magnetic ions of the crystal which are connected with the magnetic ion at the origin $\mathbf{R}_0=\mathbf{0}$ by nonvanishing exchange coupling constants $J_l \equiv J_{\mathbf{R}_l 0}$. Without loss of generality, we suppose that all the $\mathbf{R}_\mu(\mu=0,1,\ldots,n)$ are in the set L'_N defined in the second paragraph after the one containing (2.8).

Using (2.14), it is seen that, for a given $\mathbf{\Gamma}\in Z_N$, the interval

$$E_{min}^{(N)}(\mathbf{\Gamma}) \le E \le E_{max}^{(N)}(\mathbf{\Gamma}), \tag{2.15}$$

contains all the eigenvalues $\epsilon^{(N)}(\mathbf{\Gamma},\tau)$ of $h_0^{(N)}$ labeled by this $\mathbf{\Gamma}$, where

$$\begin{Bmatrix} E_{max}^{(N)}(\mathbf{\Gamma}) \\ E_{min}^{(N)}(\mathbf{\Gamma}) \end{Bmatrix} \equiv -4S \begin{Bmatrix} \min \\ \max \end{Bmatrix}_{\tau\in Z_N(\mathbf{\Gamma})} \sum_{l=1}^{n} J_l\cos(\tfrac{1}{2}\mathbf{\Gamma}\cdot\mathbf{R}_l)\cos(\tau\cdot\mathbf{R}_l). \tag{2.16}$$

Taking (2.5) and (2.9)-(2.12) into account one arrives at the following conclusions. First, $h_0^{(N)}$ and $h^{(N)}$ have nonzero matrix elements between a pair of kets $|\mathbf{\Gamma},\mathbf{R}\rangle^{(N)}$ and $|\mathbf{\Gamma}',\mathbf{R}'\rangle^{(N)}$ of $\hat{\mathcal{K}}^{(N)}$ only if $\mathbf{\Gamma}'=\mathbf{\Gamma}$, and the only nonvanishing matrix elements of $v^{(N)}$ between two such kets are those of the form

$$\langle\mathbf{\Gamma},\mathbf{R}_\mu|\, v^{(N)}|\mathbf{\Gamma},\mathbf{R}_\nu\rangle \equiv v_{\mu\nu}(\mathbf{\Gamma}), \quad \mu,\nu=0,1,\ldots,n, \tag{2.17}$$

if $\mathbf{R},\mathbf{R}'\in L'_N$ [Here and henceforth we suppress the superscript (N) occurring in the bras and kets in matrix elements.] Second, one finds that

$$\begin{aligned} &v_{0j}(\mathbf{\Gamma}) = v_{j0}(\mathbf{\Gamma}) = \sqrt{2}\eta J_j\cos(\tfrac{1}{2}\mathbf{\Gamma}\cdot\mathbf{R}_j),\\ &v_{ij}(\mathbf{\Gamma}) = -J_i\delta_{ij}, \quad i,j=1,\ldots,n,\\ &v_{00}(\mathbf{\Gamma}) = 0, \end{aligned} \tag{2.18}$$

where

$$\eta \equiv 2S\left[1-\left(1-\frac{1}{2S}\right)^{1/2}\right]. \tag{2.19}$$

In the absence of an explicit or obvious implicit indication to the contrary, the symbol $\mathbf{\Gamma}$ appearing in (2.18) and in what follows will stand for a "real" vector (one with all its components real), arbitrarily selected, and not merely for a vector in Z_N. This will be convenient in the case $N\to\infty$. In future, lower case Greek and Latin subscripts should be understood to always run over the same ranges as in (2.17) and (2.18), respectively, unless explicitly stated otherwise.

An important fact is that the interaction specified by the operator $v^{(N)}$ has both attractive and repulsive portions. More precisely, the $(n+1)$-dimensional matrix

$$v(\mathbf{\Gamma}) = \|\, v_{\mu\nu}(\mathbf{\Gamma})\|, \tag{2.20}$$

has n negative and 1 nonnegative eigenvalues for each real $\mathbf{\Gamma}$.

To establish this signature property of $v(\mathbf{\Gamma})$, let the quantities $\xi_\mu(\mu=0,1,\ldots,n)$ be arbitrary real numbers. Employing (2.18), we infer that

$$\sum_{\mu,\nu=0}^{n} v_{\mu\nu}(\mathbf{\Gamma})\,\xi_\mu\xi_\nu = 2\eta^2\left(\sum_{i=1}^{n}\cos^2(\tfrac{1}{2}\mathbf{\Gamma}\cdot\mathbf{R}_i)\right)\xi'^2_0 - \sum_{i=1}^{n} J_i\xi'^2_i, \tag{2.21}$$

where

$$\begin{aligned} &\xi'_0 = \xi_0,\\ &\xi'_i = -\sqrt{2}\cos(\tfrac{1}{2}\mathbf{\Gamma}\cdot\mathbf{R}_i)\,\xi_0 + \xi_i. \end{aligned} \tag{2.22}$$

Since the linear transformation (2.22) is nonsingular and since the diagonal quadratic form on the right hand of (2.21) has n negative coefficients, $-J_i$, and one nonnegative one, $2\eta^2\sum_{i=1}^{n}\cos^2(\tfrac{1}{2}\mathbf{\Gamma}\cdot\mathbf{R}_i)$, the asserted signature property has been proved.

Let $\mathcal{K}^{(N)}(\mathbf{\Gamma})$ be the subspace of $\mathcal{K}^{(N)}$ spanned by all the physical mixed kets $|\mathbf{\Gamma},\mathbf{R}\rangle^{(N)}$ of given $\mathbf{\Gamma}\in Z_N$. Since $H^{(N)}$ has only vanishing matrix elements between two subspaces $\mathcal{K}^{(N)}(\mathbf{\Gamma})$ and $\mathcal{K}^{(N)}(\mathbf{\Gamma}')$ with $\mathbf{\Gamma}\neq\mathbf{\Gamma}'$, we only consider the two-magnon spectrum of $H^{(N)}$ when this operator is restricted to each of the subspaces $\mathcal{K}^{(N)}(\mathbf{\Gamma})(\mathbf{\Gamma}\in Z_N)$.

At any desired $\mathbf{\Gamma}_0\in Z_N$, $H^{(N)}$ has the spectral properties stated in (1) and (2) below. In the statement of these properties and throughout this paper the following definitions obtain. Eigenvalues will always be counted according to their multiplicities. In accordance with our earlier remarks, the precise meaning of the limit $N\to\infty$ is that N tends to infinity through odd positive integral values, with all the $\mathbf{a}_j(j=1,2,3)$ and $J_k(k=1,\ldots,n)$ being kept fixed.

(1) *The number of eigenvalues of $H^{(N)}$ lying in any interval $E_1\le E\le E_2$ differs by at most $2n+3$ from the number of eigenvalues of $h_0(N)$ lying in that interval, when both of these operators are restricted to $\mathcal{K}^{(N)}(\mathbf{\Gamma}_0)$. Any two consecutive eigenvalues of the restriction of $H^{(N)}$ to this latter subspace which lie in the interval $E_{min}^{(N)}(\mathbf{\Gamma}_0)\le E\le E_{max}(\mathbf{\Gamma}_0)$ differ by $O(1/N)$ as $N\to\infty$. The latter O symbol holds in terms of a constant which is independent of $\mathbf{\Gamma}_0$ and of the pair of such consecutive eigenvalues which are considered.*[15]

(2) *The operator $H^{(N)}$, restricted to $\mathcal{K}^{(N)}(\mathbf{\Gamma}_0)$, has at most n eigenvalues (1 eigenvalue) in the interval $E< E_{min}^{(N)}(\mathbf{\Gamma}_0)[E> E_{max}^{(N)}(\mathbf{\Gamma}_0)]$.*

Property (2) gives upper bounds on the number of two-magnon "bound" states. In the spirit of the usual solid-state terminology, we shall call a state of the subspace $\mathcal{K}^{(N)}(\mathbf{\Gamma}_0)$ labeled by a given $\mathbf{\Gamma}_0\in Z_N$ a "bound" two-magnon state of total wave-number $\mathbf{\Gamma}_0$ and energy E_0 if this state is an eigenstate of $H^{(N)}$ which, when acted upon by this operator, yield the eigenvalue E_0 obeying one of the inequalities $E_0< E_{min}(\mathbf{\Gamma}_0)$ or $E_0> E_{max}(\mathbf{\Gamma}_0)$.

We now outline a proof of (1) and will prove (2) later on in this section. The first assertion (1) can be established by using two facts. One of these is that this first assertion holds if $H^{(N)}$, $\mathcal{K}^{(N)}(\mathbf{\Gamma}_0)$, and $2n+3$ are replaced by $h^{(N)}$, $\hat{\mathcal{K}}^{(N)}(\mathbf{\Gamma}_0)$, and $2(n+1)$, respectively. The symbol $\hat{\mathcal{K}}^{(N)}(\mathbf{\Gamma})$ denotes the subspace of $\hat{\mathcal{K}}^{(N)}$ spanned by all the physical and unphysical kets $|\mathbf{\Gamma},\mathbf{R}\rangle^{(N)}$ of given $\mathbf{\Gamma}\in Z_N$. One readily sees that, at each such $\mathbf{\Gamma}$, the set $\{|\mathbf{\Gamma},\mathbf{R}\rangle^{(N)}: \mathbf{R}\in L'_N\}$ of kets constitutes an orthonormal basis for the subspace $\hat{\mathcal{K}}^{(N)}(\mathbf{\Gamma})$ labeled by the $\mathbf{\Gamma}$ considered. That the first assertion (1) holds in this modified sense emerges by applying Ledermann's

theorem[16] to the $(N+1)/2 \times (N+1)/2$ matrices $\|\langle \mathbf{\Gamma}_0, \mathbf{R}' \mid h_0^{(N)} + v^{(N)} \mid \mathbf{\Gamma}_0, \mathbf{R}\rangle\|$ ($\mathbf{R}$ and $\mathbf{R}'$ range over L_N'), which are symmetric in $\mathbf{R}$ and $\mathbf{R}'$, keeping in mind the first property of $v^{(N)}$ mentioned after (2.16). The other fact is that, for each $\mathbf{\Gamma}$ of interest every eigenvalue of the restriction of $h^{(N)}$ to the subspace $\mathfrak{K}^{(N)}(\mathbf{\Gamma})$ labeled by this $\mathbf{\Gamma}$, with the possible exception of the zero eigenvalue, is an eigenvalue of the restriction of $H^{(N)}$ to the subspace $\mathfrak{K}^{(N)}(\mathbf{\Gamma})$ labeled by the same $\mathbf{\Gamma}$. One infers this with the aid of (2.10). The second assertion (1) follows from the first assertion (1) and the elementary fact that any two consecutive eigenvalues of the restriction of $h_0^{(N)}$ to each $\hat{\mathfrak{K}}^{(N)}(\mathbf{\Gamma})$ have precisely the same order property for $N \to \infty$ which we have asserted to hold for $H^{(N)}$ in (1).

Consider a real energy E_0 lying outside of the interval (2.15) labeled by $\mathbf{\Gamma} = \mathbf{\Gamma}_0 \in Z_N$. A necessary and sufficient condition for a two-magnon state of total wavenumber vector $\mathbf{\Gamma}_0$ and energy E_0 to exist is that E_0 be an eigenvalue of the restriction of $h^{(N)}$ to $\hat{\mathfrak{K}}^{(N)}(\mathbf{\Gamma}_0)$. This should be clear from the above definition of such states, the remark in the antepenultimate sentence of the previous paragraph and of the inequality $E_0 \neq 0$, this inequality resulting from the fact that the point $E = 0$ is contained in the interval (2.15) labeled by any $\mathbf{\Gamma} \in Z_N$. But as one knows from familiar reasoning, E_0 is an eigenvalue of the said restriction of $h^{(N)}$ if and only if there exists a nonzero $\varphi \in \hat{\mathfrak{K}}^{(N)}(\mathbf{\Gamma}_0)$ such that

$$\mathfrak{K}^{(N)}(E_0)\varphi = \varphi \tag{2.23}$$

obtains, where $\mathfrak{K}^{(N)}(E)$ is the Lippmann–Schwinger kernel operator

$$\mathfrak{K}^{(N)}(E) \equiv (E - h_0^{(N)})^{-1} v^{(N)}. \tag{2.24}$$

From (2.24), together with the fact that the set $\{|\mathbf{\Gamma}, \mathbf{R}\rangle^{(N)} : \mathbf{R} \in L_N'\}$ with a specified value of $\mathbf{\Gamma} \in Z_N$ constitutes an orthonormal basis for the subspace labeled by this value and with the first property of $v^{(N)}$ mentioned in the paragraph after (2.16), one readily concludes that (2.23) holds in the stated sense if and only if

$$\det[K^{(N)}(E_0, \mathbf{\Gamma}_0) - \mathbf{I}] = 0, \tag{2.25}$$

where $\mathbf{I}$ is the $(n+1) \times (n+1)$ unit matrix and

$$K^{(N)}(E, \mathbf{\Gamma}) \equiv \|\langle \mathbf{\Gamma}, \mathbf{R}_\mu \mid \mathfrak{K}^{(N)}(E) \mid \mathbf{\Gamma}, \mathbf{R}_\nu\rangle\|, \tag{2.26}$$

this definition being effective, in particular, at each pair $(E, \mathbf{\Gamma})$ which is such that $\mathbf{\Gamma} \in Z_N$ and which does not obey (2.15)

A special case of the criterion (2.25) has been employed by Boyd and Callaway[11] to determine two-magnon states and resonances. This criterion is equivalent to a generalized form of a determinantal condition of Wortis.[9] This generalization will be given later [Eq. (3.8)] in the limiting situation $N \to \infty$.

Combining (2.24) and (2.26) with the properties of $\{|\mathbf{\Gamma}, \mathbf{R}\rangle^{(N)}, \mathbf{R} \in L_N'\}$ and $v^{(N)}$ just referred to, and with (2.8), the inverse of this last equation, (2.13), (2.14), (2.17), (2.20), and the above definition of $Z_N(\mathbf{\Gamma})$, we infer that

$$K^{(N)}(E, \mathbf{\Gamma}) = -\frac{1}{2SJ} \left\| \frac{G_{\mu\nu}^{(N)}(t, \mathbf{\Gamma})}{[(1+\delta_{\mu,0})(1+\delta_{\nu,0})]^{1/2}} \right\| v(\mathbf{\Gamma})$$

$$\equiv \| K_{\mu\nu}^{(N)}(E, \mathbf{\Gamma}) \| \tag{2.27}$$

at each of the pairs $(E, \mathbf{\Gamma})$ last mentioned. The arguments E and t in (2.27) are related by the equation

$$t = -E/4SJ. \tag{2.28}$$

Equation (2.28) should be understood to hold in all succeeding equations in which both E and t appear explicitly. We define

$$J \equiv \sum_{(l)} J_l, \qquad \mathcal{J}_l \equiv J_l/J, \tag{2.29}$$

Here $\sum_{(l)}$ denotes a summation over a subset of the set of magnetic ions located at $\mathbf{R}_1, \ldots, \mathbf{R}_n$, this summation selecting exactly one magnetic ion from each of the orders of neighbors of $\mathbf{R} = 0$ which occur in this set. The $G_{\mu\nu}^{(N)}(t, \mathbf{\Gamma})$ in (2.27) are the following reduced Green's functions:

$$G_{\mu\nu}{}^{(N)}(t, \mathbf{\Gamma}) = \frac{1}{N} \sum_{\boldsymbol{\tau} \in Z_N(\mathbf{\Gamma})} \frac{\cos(\boldsymbol{\tau} \cdot \mathbf{R}_\mu) \cos(\boldsymbol{\tau} \cdot \mathbf{R}_\nu)}{\left(t - \sum_{l=1}^{u} \alpha_l(\mathbf{\Gamma}) \cos(\boldsymbol{\tau} \cdot \mathbf{R}_l)\right)}$$

$$= G_{\nu\mu}^{(N)}(t, \mathbf{\Gamma}), \tag{2.30}$$

where

$$\alpha_l(\mathbf{\Gamma}) \equiv \mathcal{J}_l \cos\tfrac{1}{2}(\mathbf{\Gamma} \cdot \mathbf{R}_l). \tag{2.31}$$

We shall regard (2.30), as well as the definitions (2.16) for $E_{min}^{(N)}(\mathbf{\Gamma})$ and $E_{max}^{(N)}(\mathbf{\Gamma})$, as holding for all real $\mathbf{\Gamma}$, where $Z_N(\mathbf{\Gamma})$ is understood to be specified at every such $\mathbf{\Gamma}$ by the same definition given previously. We now can, and shall, extend the above definition of $K^{(N)}(E, \mathbf{\Gamma})$ to each $(E, \mathbf{\Gamma})$ such that $\mathbf{\Gamma}$ is any real vector and that (2.15) is not fulfilled by requiring (2.27) to hold at each $(E, \mathbf{\Gamma})$ of this type.

From (2.30) together with (2.16), (2.29), (2.31), and an elementary argument, one can show that $\| G_{\mu\nu}^{(N)}(t, \mathbf{\Gamma}) \|$ is a positive (negative) definite matrix at those t and $\mathbf{\Gamma}$ at which $t < -E_{max}(\mathbf{\Gamma})/4SJ$ [$t > -E_{min}(\mathbf{\Gamma})/4SJ$]. From this result, the previously determined signature of the self-adjoint matrix $v(\mathbf{\Gamma})$, (2.30), and a known theorem concerning the signature of a product of two self-adjoint matrices, one of which is definite, we conclude that $K^{(N)}(E, \mathbf{\Gamma})$ has n positive (negative) eigenvalues and 1 nonpositive (1 nonnegative) eigenvalue when E and $\mathbf{\Gamma}$ are such that $E < E_{min}^{(N)}(\mathbf{\Gamma})$ [$E > E_{max}^{(N)}(\mathbf{\Gamma})$]. At any E and $\mathbf{\Gamma}$ fulfilling one of these two inequalities, the unique eigenvalue of this matrix which is nonpositive and nonnegative, respectively, in the cited cases, will be denoted by $k_0^{(N)}(E, \mathbf{\Gamma})$ and the remaining n eigenvalues by $k_i^{(N)}(E, \mathbf{\Gamma})$ $(i = 1, \ldots, n)$.

The upper bounds stated in (2) follow from these sign properties of the $k_\mu^{(N)}(E, \mathbf{\Gamma})$ and the fact that, if (2.15) is not fulfilled by a real $E = E_0$ when $\mathbf{\Gamma} = \mathbf{\Gamma}_0 \in Z_N$, a bound state of total wavenumber vector $\mathbf{\Gamma}_0$ and energy E_0 exists if and only if $K^{(N)}(E_0, \mathbf{\Gamma}_0)$ has a unit eigenvalue. This fact is obvious from the necessary and sufficient condition (2.25).

Henceforth, we shall concentrate on the case of greatest physical interest, that when $N \to \infty$.

At any real $\mathbf{\Gamma}$, we define $E_{min}(\mathbf{\Gamma})$ [$E_{max}(\mathbf{\Gamma})$] by the same Eq. (2.16) obtaining for $E_{min}^{(N)}(\mathbf{\Gamma})$ [$E_{max}^{(N)}(\mathbf{\Gamma})$], but with the maximum (minimum) taken over the paralleliped $|(\boldsymbol{\tau} \cdot \mathbf{b}_i)| \leq \pi (i = 1, 2, 3)$ (or over any region equivalent to it in the usual periodicity sense), rather than over $Z_N(\mathbf{\Gamma})$. One can show that the limit of $E_{min}^{(N)}(\mathbf{\Gamma})[E_{max}(\mathbf{\Gamma})]$ as $N \to \infty$ is $E_{min}(\mathbf{\Gamma})[E_{max}(\mathbf{\Gamma})]$. One can also prove that the interval

$$E_{min}(\mathbf{\Gamma}) \leq E \leq E_{max}(\mathbf{\Gamma}) \tag{2.32}$$

contains, for each N, the interval (2.15) pertaining to the same $\mathbf{\Gamma}$.

Let

$$K(E,\mathbf{\Gamma}) \equiv -\frac{1}{2SJ}\left\| \frac{G_{\mu\nu}(t,\mathbf{\Gamma})}{[(1+\delta_{\mu,0})(1+\delta_{\nu,0})]^{1/2}} \right\| v(\mathbf{\Gamma}) \equiv \| K_{\mu\nu}(E,\mathbf{\Gamma})\|, \tag{2.33}$$

where

$$G_{\mu\nu}(t,\mathbf{\Gamma}) \equiv \frac{1}{(2\pi)^3}\int_{-\pi}^{\pi}\int_{-\pi}^{\pi}\int_{-\pi}^{\pi} \frac{\cos(\boldsymbol{\tau}\cdot\mathbf{R}_\mu)\cos(\boldsymbol{\tau}\cdot\mathbf{R}_\nu)d\rho_1 d\rho_2 d\rho_3}{t-\sum_{i=1}^{n}\alpha_i(\mathbf{\Gamma})\cos(\boldsymbol{\tau}\cdot\mathbf{R}_i)}, \tag{2.34}$$

with

$$\rho_i \equiv (\boldsymbol{\tau}\cdot\mathbf{a}_i) \quad (i=1,2,3). \tag{2.35}$$

For a given $\mathbf{\Gamma}=\mathbf{\Gamma}_0$ and all μ and ν, it is trivially true that $G_{\mu\nu}(t,\mathbf{\Gamma}_0)$ exists as a finite number for $t<t_{min}(\mathbf{\Gamma}_0)$ and $t>t_{max}(\mathbf{\Gamma}_0)$, where

$$t_{min}(\mathbf{\Gamma}) \equiv -\frac{E_{max}(\mathbf{\Gamma})}{4SJ}, \quad t_{max}(\mathbf{\Gamma}) \equiv -\frac{E_{min}(\mathbf{\Gamma})}{4SJ}, \tag{2.36}$$

and thus that $|K_{\mu\nu}(E,\mathbf{\Gamma}_0)|<\infty$ for every μ,ν when $E<E_{min}(\mathbf{\Gamma}_0)$ or $E>E_{max}(\mathbf{\Gamma}_0)$. However, the denominator of (2.34) has zeros in the pertinent domain of integration for $\mathbf{\Gamma}=\mathbf{\Gamma}_0$ if and only if $t=t_{min}(\mathbf{\Gamma}_0)$ or $t=t_{max}(\mathbf{\Gamma}_0)$, and those zeros are well known to lead to divergent results in some cases. It is elementary that $|G_{\mu\nu}(t,\mathbf{\Gamma}_0)|<\infty$ for all μ and ν at $t=t_{min}(\mathbf{\Gamma}_0)$ or $t=t_{max}(\mathbf{\Gamma}_0)$ if and only if $|G_{00}(t,\mathbf{\Gamma}_0)|<\infty$ at these respective values of t.

It should be clear that when this property of $G_{00}(t,\mathbf{\Gamma}_0)$ holds at $t=t_{min}(\mathbf{\Gamma}_0)$ $[t=t_{max}(\mathbf{\Gamma}_0)]$, each $K_{\mu\nu}(E,\mathbf{\Gamma}_0)$ is finite for $E\overset{max}{\geq}E_{max}(\mathbf{\Gamma}_0)$ $[E\leq E_{min}(\mathbf{\Gamma}_0)]$.

Employing, in particular, Eqs. (2.27), (2.30), (2.33), (2.34), (2.35), and the generalized definition of $Z_N(\mathbf{\Gamma})$ in the second paragraph after (2.30), one finds

$$G_{\mu\nu}(t,\mathbf{\Gamma}) = \lim_{N\to\infty} G_{\mu\nu}^{(N)}(t,\mathbf{\Gamma}), \qquad K_{\mu\nu}(E,\mathbf{\Gamma}) = \lim_{N\to\infty} K_{\mu\nu}^{(N)}(E,\mathbf{\Gamma}), \tag{2.37}$$

the first of (2.37) obtaining at pairs $(t,\mathbf{\Gamma})$ such that $t<t_{min}(\mathbf{\Gamma})$ or $t>t_{max}(\mathbf{\Gamma})$, and the second at pairs $(E,\mathbf{\Gamma})$ not satisfying (2.32).

Equations of the type (2.37) expressing elementary and well-known replacements of sums by integrals in the indicated limit, provide only a purely heuristic basis for replacing, say, $K^{(N)}(E,\mathbf{\Gamma})$ by $K(E,\mathbf{\Gamma})$ in the theory of bound two-magnon states of large enough crystals, as well as for making similar substitutions in the case of other localized states in solids. Theorem 2.1, stated near the end of this section, constitutes a sound mathematical foundation for approximating the energy eigenvalues of such magnon states in the limit $N\to\infty$ (within the periodic-boundary-conditions framework) by those values of E outside of the relevant intervals (2.32) at which $K(E,\mathbf{\Gamma})$ has a unit eigenvalue for the pertinent vectors $\mathbf{\Gamma}$. Undoubtedly, similar theorems could be proved for a variety of localized states in crystals.

The following explicit formula for the latter matrix will prove useful in this paper:

$$K(E,\mathbf{\Gamma}) = \frac{1}{2S}\begin{Vmatrix} -\eta[tG_{00}(t,\mathbf{\Gamma})-1] & (1/2)[\mathcal{J}_j G_{0j}(t,\mathbf{\Gamma})-\eta\alpha_i(\mathbf{\Gamma})G_{00}(t,\mathbf{\Gamma})] \\ -\sqrt{2}\,\eta t G_{i0}(t,\mathbf{\Gamma}) & G_{ij}(t,\mathbf{\Gamma})\mathcal{J}_j-\eta G_{i0}(t,\mathbf{\Gamma})\alpha_j(\mathbf{\Gamma}) \end{Vmatrix}, \tag{2.38}$$

where, as usual, $i,j=1,\ldots,n$. This formula, which applies whenever E and $\mathbf{\Gamma}$ have the property $E\leq E_{min}(\mathbf{\Gamma})$ or $E\geq E_{max}(\mathbf{\Gamma})$ and the additional one that $|G_{00}(t,\mathbf{\Gamma})|<\infty$ at the corresponding value (2.28) of t, is a direct consequence of (2.18), (2.20), (2.29), (2.31), (2.33), and the sum rule

$$\sum_{j=1}^{n}\alpha_j(\mathbf{\Gamma})G_{\mu j}(t,\mathbf{\Gamma}) = tG_{\mu 0}(t,\mathbf{\Gamma})-\delta_{\mu,0} \tag{2.39}$$

entailed by (2.34) at the values of E and $\mathbf{\Gamma}$ just specified.

As expected, $K(E,\mathbf{\Gamma})$ has n positive (n negative) eigenvalues and 1 nonpositive (1 nonnegative) eigenvalue at each pair $(E,\mathbf{\Gamma})$ which fulfills the condition $E\leq E_{min}(\mathbf{\Gamma})$ $[E\geq E_{max}(\mathbf{\Gamma})]$ and is also such that all the $K_{\mu\nu}(E,\mathbf{\Gamma})$ are finite at this pair. Naturally, these sign properties can be proved in a way parallel to that employed to establish the signs of the eigenvalues of $K^{(N)}(E,\mathbf{\Gamma})$. At each $(E,\mathbf{\Gamma})$ not fulfilling (2.32), the unique eigenvalue of $K(E,\mathbf{\Gamma})$ which is nonpositive and nonnegative in the respective stated cases will be designated by $k_0(E,\mathbf{\Gamma})$ and the other n eigenvalues of this matrix by $k_i(E,\mathbf{\Gamma})$ $(i=1,\ldots,n)$. Moreover, at any such $(E,\mathbf{\Gamma})$, each of the $k_\mu(E,\mathbf{\Gamma})$ will be selected to be a continuous function of E. That this selection is possible can be inferred with the aid of (2.33) and the fact, easily deducible from (2.34), that the $G_{\mu\nu}(t,\mathbf{\Gamma})$ have a similar continuity property with respect to t.

The positive $k_\mu(E,\mathbf{\Gamma})$ have monotonicity properties which will play a central role in this investigation and which are of the same type as those pointed out by Rollnik[17] and Weinberg[18] for analogous eigenvalues occurring in the usual two-body problems of nonrelativistic quantum mechanics. [Properties of the same kind are possessed, of course, by the positive eigenvalues $k_\mu^{(N)}(E,\mathbf{\Gamma})$ if these are appropriately labeled.] We proceed to indicate how these properties of the $k_\mu(E,\mathbf{\Gamma})$ in question arise.

Consider a fixed, but arbitrary, subinterval of the interval $E\leq E_{min}(\mathbf{\Gamma})$ of $E\geq E_{max}(\mathbf{\Gamma})$ for a given value of $\mathbf{\Gamma}$. Then, at each $(E,\mathbf{\Gamma})$ such that $\mathbf{\Gamma}$ has the latter value and E is in the subinterval in question, every $k_\mu(E,\mathbf{\Gamma})$ is differentiable with respect to E and obeys the equation

$$\frac{\partial k_\mu(E,\mathbf{\Gamma})}{\partial E} = \rho_\mu(E,\mathbf{\Gamma})k_\mu(E,\mathbf{\Gamma}),$$

where

$$\rho_\mu(E,\mathbf{\Gamma}) \equiv \frac{\langle \mathcal{G}(E,\mathbf{\Gamma})^{-1}\partial\mathcal{G}(E,\mathbf{\Gamma})/\partial E\,\mathcal{G}(E,\mathbf{\Gamma})^{-1}\rangle_\mu}{\langle \mathcal{G}(E,\mathbf{\Gamma})^{-1}\rangle_\mu}. \tag{2.40}$$

Here $\mathcal{G}(E,\mathbf{\Gamma}) \equiv \| G_{\mu\nu}(E,\mathbf{\Gamma})/[(1+\delta_{\mu,0})(1+\delta_{\nu,0})]^{1/2}\|$, $\partial\mathcal{G}(E,\mathbf{\Gamma})/\partial E \equiv \| \partial\mathcal{G}_{\mu\nu}(E,\mathbf{\Gamma})/\partial E\|$, and $\langle A(E,\mathbf{\Gamma})\rangle_\mu \equiv \psi_\mu^+(E,\mathbf{\Gamma})A(E,\mathbf{\Gamma})\psi_\mu(E,\mathbf{\Gamma})$ where $\psi_\mu(E,\mathbf{\Gamma})$ is a column vector satisfying the eigenvalue equation $K(E,\mathbf{\Gamma})=k_\mu(E,\mathbf{\Gamma})\psi_\mu(E,\mathbf{\Gamma})$ and whose components are suitably differentiable with respect to E, and where $^+$ denotes the hermitian adjoint. We derived (2.40) by a rigorous version of Weinberg's differentiation argument.[18,19]

Every $\rho_\mu(E,\Gamma)$ is finite and positive (finite and negative) at each (E,Γ) such that $E < E_{min}(\Gamma)[E > E_{max}(\Gamma)]$. One can prove this by invoking, in particular, the easily proved positive definiteness of $\partial\mathcal{G}(E,\Gamma)/\partial E$ at those pairs (E,Γ) not obeying (2.32) and the definiteness properties of $\mathcal{G}(E,\Gamma)$ at each (E,Γ) of this type.

One can now conclude the following about the eigenvalues of $K(E,\Gamma)$ which are nonnegative on the pertinent energy intervals, in view of the above continuity property of the $k_\mu(E,\Gamma)$ with respect to E, the circumstance that (2.40) obtains within the stated possible exceptions on each subinterval of the indicated type, and the properties of the $\rho_\mu(E,\Gamma)$ just mentioned. At any chosen $\Gamma = \Gamma_0$, each $k_i(E,\Gamma_0)$ is a strictly increasing function of E at all those $E \le E_{min}(\Gamma_0)$ at which all the $K_{\mu\nu}(E,\Gamma_0)$ are finite, and $k_0(E,\Gamma_0)$ is either strictly decreasing in E or vanishes at all those $E \ge E_{max}(\Gamma_0)$ at which this finiteness condition obtains for each μ and ν.

Before stating Theorem 2.1, some definitions are in order. The $k_\mu^{(N)}(E,\Gamma)$ referred to in this theorem should be understood to have been chosen as jointly continuous in E and all the $(\Gamma\cdot\mathbf{a}_k)(k = 1,2,3)$ at each (E,Γ) at which (2.32) does not hold. An argument analogous to one outlined previously shows that this choice is possible. In addition, the $k_\mu^{(N)}(E,\Gamma)$ are assumed to have been selected in such a way that for each μ the limit of $k_\mu^{(N)}(E,\Gamma)$ for $N \to \infty$ exists at every (E,Γ) of the type just mentioned. It is readily shown that these two requirements can be fulfilled simultaneously. The $k_\mu(E,\Gamma)$ of Theorem 2.1 are defined as follows in terms of a given set of eigenvalues $k_\mu^{(N)}(E,\Gamma)$ satisfying both of these conditions:

$$k_\mu(E,\Gamma) = \lim_{N\to\infty} k_\mu^{(N)}(E,\Gamma). \tag{2.41}$$

One can show that the definition (2.41), which is very natural in the context of the said theorem, is consistent with our previous definition of the $k_\mu(E,\Gamma)$.

Theorem 2.1: Consider a fixed $\mu = 0,1,\ldots n$ and a fixed $\epsilon > 0$. Denote by S_ϵ the set of all pairs (E,Γ) for which $\Gamma \in Z$ and either $E \le E_{min}(\Gamma) - \epsilon$ or $E \ge E_{max}(\Gamma) + \epsilon$. Let $\hat{\Gamma}$ be a vector in Z at which the equation

$$k_\mu(\hat{E},\hat{\Gamma}) = 1$$

has the (unique) solution $\hat{E}$ such that $(\hat{E},\hat{\Gamma}) \in S_\epsilon$. At each odd positive integer N, let Γ_N, be a vector in Z_N, these vectors converging to Γ:

$$\lim_{N\to\infty} |\hat{\Gamma} - \Gamma_N| = 0.$$

At each such N and every real Γ, denote by $E_N(\Gamma)$ the solution (unique if it exists) of

$$k_\mu^{(N)}(E,\Gamma) = 1,$$

such that $E_N(\Gamma)$ lies outside of the interval (2.25) pertaining to the Γ in question. Then $E_N(\Gamma_N)$ exists at each $N \ge N_0$ of this type, where $N_0 = N_0(\epsilon)$ is a positive integer independent of N, $\hat{\Gamma}$, and the set of vectors Γ_N considered, and

$$\lim_{N\to\infty} E_N(\Gamma_N) = \hat{E},$$

this limit being attained uniformly with respect to $\hat{\Gamma}$.

Our proof of this theorem will be omitted because of its length.[20] Notice that a set of such vectors Γ_N exists for every $\Gamma \in Z$ and that the uniqueness properties in the theorem are trivial consequences of the pertinent monotonicity properties of the k_μ and $k_\mu^{(N)}$.

There is a clear heuristic reason for introducing the arbitrarily small positive ϵ, which remains fixed as $N \to \infty$. Indeed, one cannot expect that the standard replacement of sums by integrals will be accurate, even for large N, unless the energy eigenvalues of the bound states of interest lie at distances from the edges of the relevant band (2.32) which are large compared with the spacing of the pertinent eigenvalues $\epsilon^{(N)}(\Gamma,\tau)$.

3. RIGOROUS UPPER AND LOWER BOUNDS ON THE NUMBER OF BOUND STATES

Having laid the foundations in Sec. 2, we are now prepared to discuss our techniques, mainly of the trace-inequality type, for the study of two-magnon bound states in Heisenberg ferromagnets in the limiting case $N \to \infty$, the only case which will be considered henceforth.

In dealing with the case $E = E_{min}(\Gamma)[E = E_{max}(\Gamma)]$, we shall always suppose from now on that Γ is such that $|G_{00}(t,\Gamma)| < \infty$ at $t = t_{max}(\Gamma)[t = t_{min}(\Gamma)]$. This finiteness condition will make all of our subsequent trace equalities and inequalities meaningful in these two cases. The first, less restrictive, of the two definitions of the $k_\mu(E,\Gamma)$ given in Sec. 2 applies in all of our subsequent discussions.

Whenever we speak in future of the existence or nonexistence of bound states with energies E in a specified range and with (or at) given values of $\Gamma \in Z$, or use obviously equivalent terms, we shall naturally mean, in the spirit of Theorem 2.1, that the appropriate matrix $K(E,\Gamma)$ has a unit eigenvalue or no unit eigenvalue, respectively, at the values of E and Γ not obeying (2.32) which are considered.

For each $\Gamma_0 \in Z$ we denote by $N_1(E_0,\Gamma_0)[N_2(E_0,\Gamma_0)]$ the total number of times that the $n+1$ eigenvalues $k_\mu(E,\Gamma_0)$ attain the value unity in any given interval $E < E_0 \le E_{min}(\Gamma_0)[E > E_0 \ge E_{max}(\Gamma_0)]$.

The following two theorems, the first of which is analogous to a result obtained by Schwinger[21] in a two-body nonrelativistic context, provide upper bounds on the number of bound states. In the statement of these and the other four theorems, Γ_0 will denote an arbitrary fixed vector in Z.

Theorem 3.1: One has

$$N_{1,2}(E_0,\Gamma_0) < \mathrm{Tr}K^{2m}(E_0,\Gamma_0) \tag{3.1}$$

for all $m = 1,2,\ldots$ when $E_0 \le E_{min}(\Gamma_0)$ and $E_0 \ge E_{max}(\Gamma_0)$, respectively.

Proof: Because of the stated monotonicity and continuity of the $k_\mu(E,\Gamma_0)$ as functions of E, one infers that $N_1(E_0,\Gamma_0)$ and $N_2(E_0,\Gamma_0)$ are equal to the total number of $k_\mu(E_0,\Gamma_0)$ which are larger than unity when $E_0 \le E_{min}(\Gamma_0)$ and $E_0 \ge E_{max}(\Gamma_0)$, respectively. Hence, when E_0 and Γ_0 obey these respective inequalities,

$$N_{1,2}(E_0,\Gamma_0) < \sum_\alpha k_\alpha^r(E_0,\Gamma_0), \tag{3.2}$$

for any nonnegative integer r, α running over all the positive eigenvalues $k_\alpha(E_0,\Gamma_0)$ of $K(E_0,\Gamma_0)$. Therefore,

$$N_{1,2}(E_0,\Gamma_0) < \sum_{\nu=0}^{n} k_\nu^{2m}(E_0,\Gamma_0) = \mathrm{Tr}K^{2m}(E_0,\Gamma_0)$$

in these two respective cases for any positive integer m.

Theorem 3.2: Let p and q be positive integers such that p is odd and q is even, and let $E_0 \leq E_{min}(\Gamma_0)$. Then

$$N_1(E_0,\Gamma_0) < \lambda_{pq}(E_0,\Gamma_0) \equiv \mathrm{Tr}K^p(E_0,\Gamma_0) + [\mathrm{Tr}K^q(E_0,\Gamma_0)]^{p/q}. \quad (3.3a)$$

Moreover if p and q possess these properties and are such that $p > q$, then

$$N_{1,2}(E_0,\Gamma_0) < \tfrac{1}{2}\lambda_{pq}(E_0,\Gamma_0) \quad (3.3b)$$

in the respective intervals $E_0 \leq E_{min}(\Gamma_0)$ and $E_0 \geq E_{max}(\Gamma_0)$.

Proof: Let E_0, Γ_0, p, and q be selected to satisfy the conditions of the first sentence of the theorem. From the inequalities $k_0(E_0,\Gamma_0) \leq 0$ and $k_i(E_0,\Gamma_0) > 0$ which then hold, and from the assumed properties of p and q, one finds

$$\lambda_{pq}(E_0,\Gamma_0) = \sum_{i=1}^{n} k_i^p(E_0,\Gamma_0) + \left[\left(k_0^q(E_0,\Gamma_0) + \sum_{i=1}^{n} k_i^q(E_0,\Gamma_0)\right)^{p/q} - |k_0^p(E_0,\Gamma_0)|\right].$$

Noticing that the quantity inside the square brackets in this equation is positive for such a choice of parameters and invoking (3.2), one concludes that (3.3a) obtains under the stated circumstances.

To prove the remaining portion of the theorem, we proceed as follows. If E_0, Γ_0, p, and q are chosen in accordance with the third sentence of the theorem, so that, in particular, $p > q$, we infer with the aid of Jensen's inequality[22] that

$$[\mathrm{Tr}K^q(E_0,\Gamma_0)]^{p/q} = \left(\sum_{\mu=0}^{n} k_\mu^q(E_0,\Gamma_0)\right)^{p/q} \geq \sum_{\mu=0}^{n} k_\mu(E_0,\Gamma_0)|.$$

Whence

$$\lambda_{pq}(E_0,\Gamma_0) > \sum_{\mu=0}^{n} k_\mu^p(E_0,\Gamma_0) + \sum_{\mu=0}^{n} |k_\mu^p(E_0,\Gamma_0)| = 2\sum_{\alpha} k_\alpha^p(E_0,\Gamma_0)$$

for this selection, and thus, recalling (3.2), we see that the desired inequality (3.3b) holds under the conditions of interest.

For the sake of convenient reference in subsequent discussions, we state Theorems 3.3 and 3.4, which are immediate consequences of Theorems 3.1 and 3.2 respectively.

Theorem 3.3: When $E_0 \leq E_{min}(\Gamma_0)$ $[E_0 \geq E_{max}(\Gamma_0)]$, no bound states with $\Gamma = \Gamma_0$ and $E < E_0 \leq E_{min}(\Gamma_0)$ $[E > E_0 \geq E_{max}(\Gamma_0)]$ exist if

$$\mathrm{Tr}K^{2m}(E_0,\Gamma_0) < 1 \quad (3.4)$$

for some $m = 1, 2, \cdots$.

Theorem 3.4: No bound states with $\Gamma = \Gamma_0$ and $E < E_0 \leq E_{min}(\Gamma_0)$ exist if there is a pair of positive integers p and q, odd and even, respectively, and such that

$$\lambda_{pq}(E_0,\Gamma_0) \leq 1. \quad (3.5a)$$

Moreover, a sufficient condition for the nonexistence of bound states with $\Gamma = \Gamma_0$ and either $E < E_0 \leq E_{min}(\Gamma_0)$ or $E > E_0 \geq E_{max}(\Gamma_0)$ is that there exist two positive integers p and q with these properties and the added one $p > q$, such that

$$\lambda_{pq}(E_0,\Gamma_0) \leq 2. \quad (3.5b)$$

While the simple trace inequality (3.4) evidently fails in every case when $E_0 \leq E_{min}(\Gamma_0)$ and $E_0 \geq E_{max}(\Gamma_0)$ if one of the negative eigenvalues $k_\mu(E_0,\Gamma_0)$ is larger than unity in absolute value, the trace inequalities (3.5a) and (3.5b) are obeyed in certain cases of this kind, some of which will be dealt with in Sec. 5 of this paper. Indeed, it can be shown that if there are no bound states with $\Gamma = \Gamma_0$ in the energy region $E < E_0 \leq E_{min}(\Gamma_0)$, there exist positive integers p and $q < p$, odd and even respectively, such that $\lambda_{pq}(E_0,\Gamma_0)$ is smaller than any preassigned positive number.

Let us now turn to a criterion for the existence of bound states.

Theorem 3.5: For $E_0 \leq E_{min}(\Gamma_0)[E_0 \geq E_{max}(\Gamma_0)]$, a sufficient condition for the existence of at least one bound state with $\Gamma = \Gamma_0$ and $E < E_0 (E > E_0)$ is that there be an odd positive integer p and an even positive integer $q < p$ such that

$$\mathrm{Tr}K^p(E_0,\Gamma_0) > \mathrm{Tr}K^q(E_0,\Gamma_0). \quad (3.6)$$

Proof: Let $E_0 \leq E_{min}(\Gamma_0)$ or $E_0 \geq E_{max}(\Gamma_0)$, and let E_0, Γ_0, p and q satisfy the conditions asserted to be sufficient in the theorem. Denoting by $\sum_\beta$ a summation over all the nonpositive eigenvalues $k_\beta(E_0,\Gamma_0)$ of $K(E_0,\Gamma_0)$, we then find

$$\begin{aligned} 0 &< \mathrm{Tr}K^p(E_0,\Gamma_0) - \mathrm{Tr}K^q(E_0,\Gamma_0) \\ &= \sum_\alpha k_\alpha^q(E_0,\Gamma_0)[k_\alpha^{p/q}(E_0,\Gamma_0) - 1] \\ &\quad - \sum_\beta k_\beta^q(E_0,\Gamma_0)[|k_\beta^{p/q}(E_0,\Gamma_0)| + 1] \\ &\leq \sum_\alpha k_\alpha^q(E_0,\Gamma_0)[k_\alpha^{p/q}(E_0,\Gamma_0) - 1]. \end{aligned}$$

This result entails, since $p > q$ in the present case, that some positive eigenvalue $k_\alpha(E_0,\Gamma_0)$ is greater than unity. We thus conclude that $N_{1,2}(E_0,\Gamma_0) \geq 1$ in the respective cases $E_0 \leq E_{min}(\Gamma_0)$ and $E_0 \geq E_{max}(\Gamma_0)$ when (3.6) holds in the cases in question.

Theorems 3.3, 3.4 and 3.5 have been found to be particularly useful in the investigation of the existence of bound states at energies below $E_{min}(\Gamma)$, for various ranges of Γ considered in the detailed calculations of Secs. 4 and 5. The next theorem has proved to be valuable in showing that no bound states occur at energies above $E_{max}(\Gamma)$ for ranges of this vector dealt with in those calculations.

Theorem 3.6: Let E_1 and Γ_0 be such that $E_1 \geq E_{max}(\Gamma_0)$ and that

$$\left(\max_{1\leq i\leq n} \mathcal{J}_i\right)\left(\sum_{i=1}^{n} \beta_i^2(\Gamma_0)/8S\right)|G_{00}(t_1,\Gamma_0)| \leq 1, \quad (3.7)$$

where $t_1 \equiv -E_1/4SJ$ and $\beta_i(\Gamma) \equiv \cos\frac{1}{2}(\Gamma\cdot\mathbf{R}_i)$. Then there are no bound states with $\Gamma = \Gamma_0$ in the energy range $E > E_1 \geq E_{max}(\Gamma_0)$.

Proof: In this proof we shall use the fact that a necessary and sufficient condition that $K(E,\Gamma)$ have at least one unit eigenvalue at a prescribed pair (E,Γ) such that $E \leq E_{min}(\Gamma)$ or $E \geq E_{max}(\Gamma)$ is that

$$\det \boldsymbol{M}(-E/4SJ,\Gamma) = 0 \quad (3.8)$$

at this (E,Γ), where $M(t,\Gamma)$ is the $n \times n$ matrix with elements

$$M_{ij}(t,\Gamma) \equiv -G_{ij}(t,\Gamma) + G_{i0}(t,\Gamma)\beta_j(\Gamma) + 2S\delta_{i,j}\mathcal{J}_j^{-1}. \tag{3.9}$$

This fact can be easily derived, via simple algebraic manipulations, from (2.38), (2.39), and the circumstance that $K(E,\Gamma)$ has this eigenvalue property at the (E,Γ) in question if and only if

$$\det[K(E,\Gamma) - \mathbf{I}] = 0$$

at this pair. In the derivation of (3.8) just alluded to it is assumed that $E \neq 0$, a condition automatically satisfied at all values of E and Γ obeying the finiteness condition stated earlier in this section

Let (t_1,Γ_0) be a pair satisfying the conditions of the theorem, and let $t_2 \equiv -E_2/4SJ$, with $E_2 > E_1 \geq E_{max}(\Gamma_0)$. Then $|G_{00}(t_2,\Gamma_0)| < |G_{00}(t_1,\Gamma_0)|$, as follows from (2.34) by means of elementary considerations. Since (3.7) has been assumed to hold at (t_1,Γ_0), we thus conclude that

$$\left(\max_{1\leq i\leq n}\mathcal{J}_i\right)\left(\sum_{i=1}\beta_i^2(\Gamma_0)/8S\right)|G_{00}(t_2,\Gamma_0)| < 1. \tag{3.10}$$

To prove the theorem, we shall show that (3.10) entails that the symmetric part of the real matrix $M(t_2,\Gamma_0)$, in the unique decomposition of $M(t_2,\Gamma_0)$ into symmetric and antisymmetric matrices, is positive definite. This property of the latter symmetric part implies[23] that the real parts of all the eigenvalues of $M(t_2,\Gamma_0)$ are positive, and hence that (3.8) cannot hold at $E = E_2$ when $\Gamma = \Gamma_0$, from which the theorem obviously follows.

Denote by $\mathfrak{M}(t,\Gamma)$ the $n \times n$ matrix with elements

$$\mathfrak{M}_{ij}(t,\Gamma) \equiv \frac{1}{(2\pi)^3}\int_{-\pi}^{\pi}\int_{-\pi}^{\pi}\int_{-\pi}^{\pi} \frac{[\cos(\boldsymbol{\tau}\cdot\mathbf{R}_i) - \frac{1}{2}\beta_i(\Gamma)][\cos(\boldsymbol{\tau}\cdot\mathbf{R}_j) - \frac{1}{2}\beta_j(\Gamma)]}{t - \sum_{l=1}^{n}\alpha_l(\Gamma)\cos(\boldsymbol{\tau}\cdot\mathbf{R}_l)}\, d\rho_1 d\rho_2 d\rho_3. \tag{3.11}$$

An argument parallel to one alluded to earlier shows that $\mathfrak{M}(t_2,\Gamma_0)$ is positive definite.

From (3.10) and (3.11), the symmetric part of $M(t,\Gamma)$ is

$$M_S(t,\Gamma) = \mathfrak{M}(t,\Gamma) + \mathfrak{N}(t,\Gamma), \tag{3.12}$$

where

$$\mathfrak{N}(t,\Gamma) \equiv 2S\mathcal{J}^{-1} + \tfrac{1}{4}G_{00}(t,\Gamma)\beta(\Gamma)\mathcal{T}\beta(\Gamma). \tag{3.13}$$

Here $\beta(\Gamma)$ and $\mathcal{J}$ are the $n \times n$ diagonal matrices $[\beta_i(\Gamma)\delta_{ij}]$ and $[\mathcal{J}_i\delta_{ij}]$, respectively, $\mathcal{J}$ being obviously positive definite, and $\mathcal{T}$ is the positive semidefinite matrix all of whose elements are equal to unity. As we shall prove shortly, $\mathfrak{N}(t_2,\Gamma_0)$ is positive definite. Combining this fact with (3.12) and the positive definiteness of $\mathfrak{M}(t_2,\Gamma_0)$, the promised result that $M_S(t_2,\Gamma_0)$ has this same property follows.

A particularly suggestive notation with which to prove the positive definiteness of $\mathfrak{N}(t_2,\Gamma_0)$ is the inner product notation $(\varphi,\psi) \equiv \sum_{j=1}^{n}\varphi_j\psi_j$, where $\varphi = [\varphi_i]$ and $\psi = [\psi_i]$ $(i = 1,\ldots,n)$ are column vectors with n real components. Using in particular (3.13) the familiar inequality

$$(\varphi, \mathcal{C}\varphi) \leq \gamma(\varphi,\varphi),$$

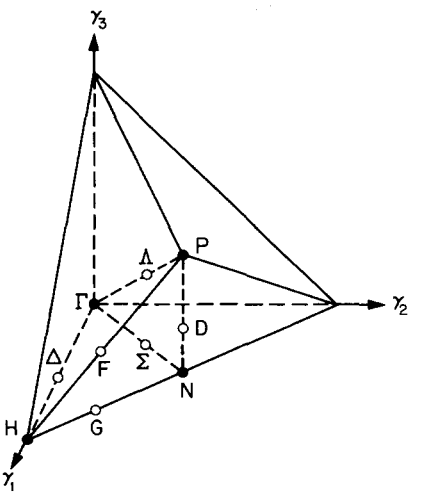


FIG 1. One eighth of the Brillouin zone of the bcc structure, showing the pertinent points and lines of symmetry.

where $\mathcal{C}$ is an $n \times n$ symmetric matrix and γ is its maximum eigenvalue, the fact that the maximum eigenvalue of $\beta(\Gamma_0)\mathcal{T}\beta(\Gamma_0)$ is equal to $\sum_{i=1}^{n}\beta_i^2(\Gamma_0)$, the inequality $G_{00}(t_2,\Gamma_0) < 0$, easily obtainable from (2.34), and (3.10) we readily conclude that

$$\begin{aligned}(\varphi,\mathfrak{N}(t_2,\Gamma_0)\varphi) &= 2S(\varphi,\mathcal{J}^{-1}\varphi) - \tfrac{1}{4}|G_{00}(t_2,\Gamma_0)|(\varphi,\beta(\Gamma_0)\mathcal{T}\beta(\Gamma_0)\varphi)| \\ &\geq \left[2S\left(1/\max_{1\leq i\leq n}\mathcal{J}_i\right) - \tfrac{1}{4}|G_{00}(t_2,\Gamma_0)|\sum_{i=1}^{n}\beta_i^2(\Gamma_0)\right](\varphi,\varphi) > 0,\end{aligned}$$

if $(\varphi,\varphi) > 0$, i.e., that $\mathfrak{N}(t_2,\Gamma_0)$ is positive definite. Hence we have completed proving the theorem.

4. BCC STRUCTURE WITH NN INTERACTIONS: Γ IN THE [100] DIRECTION

Henceforth we shall be mainly interested in illustrating the techniques of Sec. 3 for the example when the magnetic ions constitute a bcc lattice and are only coupled by NN interactions. An eighth of the first Brillouin zone Z_{bcc} of the bcc structure, including a complete set of inequivalent special points and special lines of this zone, is shown in Fig. 1 for future reference.[10] In this figure, the arrows refer to the respective directions of three mutually perpendicular unit vectors $\mathbf{i}_p (p = 1,2,3)$, each such $\mathbf{i}_p$ pointing of course along one of the cubic axes, and $\gamma_p \equiv (\Gamma\cdot\mathbf{i}_p)$ for each of these unit vectors.

Before studying the case when Γ is a vector of Z_{bcc} parallel to the [100] direction, we give some formulas which will prove useful in the sequel.

The basis vectors $\mathbf{a}_p (p = 1,2,3)$ of a bcc lattice and the vectors $\mathbf{R}_l (l = 1,2,3,4)$ connecting a given site of such a lattice to half of its nearest-neighbor sites will be chosen as specified by the equations below:

$$\begin{aligned}&\mathbf{a}_p = (a/2)(-\mathbf{i}_p + \mathbf{i}_q + \mathbf{i}_r), \\ &\mathbf{R}_l = \mathbf{a}_l \quad (l = 1,2,3), \qquad \mathbf{R}_4 = \mathbf{a}_1 + \mathbf{a}_2 + \mathbf{a}_3,\end{aligned} \tag{4.1}$$

where p,q,r are a cyclic permutation of 1, 2, 3 and a is the length of the edge of the elementary cube of the bcc lattice. We shall use units of length such that a is unity.

When NN coupling alone is present, the following equation holds at each Γ for the sc and the bcc structures:

TABLE I. Formulas for α_l and $\hat{t}$ for a complete set of inequivalent special points and special lines of the first Brillouin zone of the bcc structure for the case of purely NN interactions.

		($\alpha = 1$)	
Γ	Special Points and Special Lines	α_l	$\hat{t}$
$(\gamma, 0, 0)$	Γ: $\gamma = 0$ H: $\gamma = 2\pi$, Δ: $0 < \gamma < 2\pi$	$\alpha_l = \cos(\gamma/4) = \alpha$, $l = 1, 2, 3, 4$	4α
(γ, γ, γ)	P: $\gamma = \pi$, Λ: $0 < \gamma < \pi$	$\alpha_1 = \alpha_2 = \alpha_3$ $= \cos(\gamma/4)$, $\alpha_4 = \cos(3\gamma/4)$	$4\alpha_1^3$ if $0 < \gamma \leq \gamma_c$;[a] $\dfrac{8\alpha_1(1-\alpha_1^2)^{3/2}}{(3-4\alpha_1^2)^{1/2}}$ if $\gamma_c < \gamma \leq \pi$
$(\gamma, \gamma, 0)$	Σ: $0 < \gamma < \pi$	$\alpha_1 = \alpha_2 = 1$, $\alpha_3 = -\alpha_4 =$ $\cos(\gamma/2)$	$2(1 + \alpha_3)$
(π, π, γ)	D: $0 < \gamma < \pi$	$\alpha_1 = \alpha_2 =$ $\cos(\gamma/4)$, $\alpha_3 = -\alpha_4 =$ $\sin(\gamma/4)$	2
$(2\pi - \gamma, \gamma, \gamma)$	F: $0 < \gamma < \pi$	$\alpha_1 = \sin(3\gamma/4)$, $\alpha_2 = \alpha_3 = -\alpha_4$ $= \sin(\gamma/4)$	$\dfrac{8\alpha_2(1-\alpha_2^2)^{3/2}}{(3-4\alpha_2^2)^{1/2}}$
$(2\pi - \gamma, \gamma, 0)$	N: $\gamma = \pi$, G: $0 < \gamma < \pi$	$\alpha_1 = \alpha_2 = \sin(\gamma/2)$ $\alpha_3 = \alpha_4 = 0$	$2\alpha_1$

[a] γ_c is the unique solution of $\cos(\gamma/4) = \sqrt{2/3}$ in the interval $0 < \gamma < \pi$.

$$E_{min}(\Gamma) = -E_{max}(\Gamma).$$

From this equation, (2.28), and (2.36), we may write

$$t_{max}(\Gamma) = -t_{min}(\Gamma) \equiv \hat{t}(\Gamma). \tag{4.2}$$

Until further notice, we shall concentrate exclusively on bcc ferromagnets with solely NN coupling. Table I gives the values of Γ corresponding to the points and lines of symmetry of Z_{bcc} exhibited in Fig. 1, as well as formulas for $\hat{t}(\Gamma)$ and $\alpha_l(\Gamma)$ $(l = 1, 2, 3, 4)$ appropriate to the values in question.

In this table and in what follows, a vector $\Gamma = \sum_{p=1}^{3} \gamma_p \mathbf{i}_p$ is denoted by $(\gamma_1, \gamma_2, \gamma_3)$.

From Eqs. (2.34), (2.35), and (4.1), we arrive at the integral representation below in the case of ferromagnets of the latter structure and coupling:

$$G_{\mu\nu}(t, \Gamma) = \frac{1}{(2\pi)^3} \int_{-\pi}^{\pi}\int_{-\pi}^{\pi}\int_{-\pi}^{\pi} \frac{\cos\rho_\mu \cos\rho_\nu d\rho_1 d\rho_2 d\rho_3}{t - \sum_{l=1}^{4} \alpha_l(\Gamma)\cos\rho_l}, \tag{4.3}$$

$$\rho_0 \equiv 0, \quad \rho_4 \equiv \rho_1 + \rho_2 + \rho_3.$$

In accordance with our earlier convention and the fact that $n = 4$ for the ferromagnets being considered, the indices μ, ν range from 0 to 4 in (4.3) and elsewhere in this and the succeeding section, and the indices i, j range from 1 to 4 in these two sections.

With the aid of (4.3), one concludes that

$$\begin{aligned} G_{00}(-t, \Gamma) &= -G_{00}(t, \Gamma), \\ G_{i0}(-t, \Gamma) &= G_{i0}(t, \Gamma), \\ G_{ij}(-t, \Gamma) &= -G_{ij}(t, \Gamma), \end{aligned} \tag{4.4}$$

at those real t and Γ which are such that both $|t| \leq \hat{t}(\Gamma)$ and $|G_{00}(t, \Gamma)| < \infty$.

For the bcc example under discussion, we are interested primarily in the question of whether bound states exist at all at energies below $E_{min}(\Gamma)$ or above $E_{max}(\Gamma)$, although we feel that much more than this could be obtained in many cases by a sagacious application of suitable theorems of Sec. 3. Because of (4.2), (4.4), and the fact that in the relevant equations [for example in (2.33] t and E are related by (2.28), we may limit ourselves, as far as this existence question is concerned, to the evaluation of the functions $G_{\mu\nu}(t, \Gamma)$ at $t = \hat{t}(\Gamma)$ when Γ is such that $|G_{00}(\hat{t}(\Gamma), \Gamma)| < \infty$.

Let us turn to the situation when Γ is a vector of Z_{bcc} of the form $(\gamma, 0, 0)$, with $0 \leq \gamma \leq 2\pi$.

We begin by calculating the pertinent quantities $G_{\mu\nu}(\hat{t}, \Gamma) \equiv G_{\mu\nu}(\hat{t}(\Gamma), \Gamma)$ at such values of Γ. The calculations are greatly simplified by the fact, exhibited in Table I, that all the $\alpha_l(\Gamma)$ in (4.3) are equal to the same number $\alpha(\Gamma)$ at each of these Γ. If Γ coincides with the central point Γ of Z_{bcc} or is a point of the special line Δ, then $0 < \alpha(\Gamma) \leq 1$; and if it is coincident with the special point H on the zone boundary, then $\alpha(\Gamma) = 0$.

We shall denote the line segment consisting of the point Γ and of the points of Δ by the symbol Δ'.

From (4.3) and the vanishing of $\alpha(\Gamma)$ when Γ is at H, it is seen that at this value of Γ

$$\begin{aligned} G_{\mu 0}(t, \Gamma) &= (1/t)\,\delta_{\mu 0}, \\ G_{\mu j}(t, \Gamma) &= (1/2t)\,\delta_{\mu j}. \end{aligned} \tag{4.5}$$

Because of the singular behavior of the $G_{\mu\nu}(t, \Gamma)$ in (4.5) as $t \to 0$, it will prove convenient to handle this trivial subcase separately. From the property $\alpha(\Gamma) \neq 0$ holding when $\Gamma \in \Delta'$ and by the application of familiar transformation procedures to the integrals (4.3), we see that at every such Γ the $G_{\mu\nu}(\hat{t}, \Gamma)$ are finite and that each of them can be expressed as a linear combination of the following Γ-independent integrals:

$$\begin{aligned} I_0 &\equiv \frac{1}{\pi^3}\int_0^\pi\int_0^\pi\int_0^\pi \frac{dudvdw}{1 - \cos u \cos v \cos w}, \\ I_1 &\equiv \frac{1}{\pi^3}\int_0^\pi\int_0^\pi\int_0^\pi \frac{dudvdw\,\cos^2 u}{1 - \cos u \cos v \cos w}, \\ I_2 &= \frac{1}{\pi^3}\int_0^\pi\int_0^\pi\int_0^\pi \frac{dxdydz\,\cos^2 u \cos^2 v}{1 - \cos u \cos v \cos w}, \end{aligned} \tag{4.6}$$

Explicitly, we have at each of the latter Γ:

$$\left.\begin{aligned} &G_{\mu\nu}(\hat{t}, \Gamma) = \frac{1}{4\alpha(\Gamma)}\,\mathcal{G}_{\mu\nu}, \\ &\text{where} \\ &\mathcal{G}_{00} = I_0, \\ &\mathcal{G}_{0i} = \mathcal{G}_{i0} = I_0 - 1, \\ &\mathcal{G}_{ji} = 4(I_0 - 1) + 3(I_1 - 2I_2), \\ &\mathcal{G}_{ij} = \mathcal{G}_{ji} = 2I_2 - I_1, i \neq j. \end{aligned}\right\} \tag{4.7}$$

The following closed-form expressions for I_0, I_1, and I_2 in terms of the complete elliptic integral K of the indicated modulus hold:

$$I_0 = \frac{4}{\pi^2} K^2\left(\frac{1}{\sqrt{2}}\right) = 1.3932093,$$

$$I_1 = \frac{2}{\pi^2}\left[K^2\left(\frac{1}{\sqrt{2}}\right) + \frac{\pi^2}{4K^2\left(\frac{1}{\sqrt{2}}\right)}\right] = 0.8420526, \qquad (4.8)$$

$$I_2 = \frac{2}{\pi^2}\left[K^2\left(\frac{1}{\sqrt{2}}\right) - \frac{\pi^2}{4K^2\left(\frac{1}{\sqrt{2}}\right)}\right] = 0.5511514.$$

As is well known, Watson[24] derived a closed form result for I_0 equivalent to the one shown in (4.8). The analytical expressions for I_1 and I_2 in (4.8) were proved recently by Joyce.[25]

After these preparations, let us investigate the occurrence of bound states for ferromagnets of bcc structure with only NN interactions when $\mathbf{\Gamma}$ is at Γ or on Δ'.

At any such $\mathbf{\Gamma}$, no bound states with energies in the range $E > E_{max}(\mathbf{\Gamma})$ exist for such ferromagnets at any $S \geq \frac{1}{2}$.[26] At the unique vector $\mathbf{\Gamma}$ of this type at which $\alpha(\mathbf{\Gamma}) = 0$, this assertion follows trivially by direct calculation or by applying Theorem 3.6 at those E_1 that are greater than the pertinent energy $E_{max}(\mathbf{\Gamma})$, keeping in mind that $\beta_i(\mathbf{\Gamma}) = 0$ at this unique vector for the latter ferromagnets. For any such ferromagnet with a given $S \geq \frac{1}{2}$, we conclude that no bound states with energies above $E_{max}(\mathbf{\Gamma})$ are possible at any of the other vectors $\mathbf{\Gamma}$ of the type to which we have just alluded, i.e., when $\mathbf{\Gamma}$ coincides with Γ or lies on Δ, provided that the inequality

$$I_0 \leq 8S \qquad (4.9)$$

is satisfied at the value of S considered. This conclusion follows by applying Theorem 3.6 for E_1 equal to the relevant values of $E_{max}(\mathbf{\Gamma})$, using the facts that $0 < \beta_i(\mathbf{\Gamma}) \leq 1$ for the ferromagnets under discussion when $\mathbf{\Gamma}$ is at Γ or on Δ and that $n = 4$ for these ferromagnets, and by employing (2.36), (4.2), (4.4), and (4.7). That (4.9) obtains for $S \geq \frac{1}{2}$ is obvious from (4.8).

For the pertinent ferromagnets, we proceed to establish rigorously a threshold for the appearance of bound states with energies $E < E_{min}(\mathbf{\Gamma})$ for the range of $\mathbf{\Gamma}$ of interest, by employing exact expressions for the eigenvalues of $K(E, \mathbf{\Gamma})$ at $E = E_{min}(\mathbf{\Gamma})$. We shall then compare this exact threshold formula with rigorous upper and lower bounds to the threshold in question obtained by three of the trace-inequality methods of Sec. 3.

When specialized to the bcc structure with NN coupling only, (2.38) leads to the results below if $\mathbf{\Gamma}$ is at Γ or at Δ', as is inferred by employing (4.7) and the facts that $\alpha_j(\mathbf{\Gamma}) = \alpha(\mathbf{\Gamma})$ and that $\mathcal{J}_i = 1$ for the situation under consideration:

$$\begin{aligned}
K_{00}(E, \mathbf{\Gamma}) &= -\eta[tG_{00}(t, \mathbf{\Gamma}) - 1], \\
K_{0i}(E, \mathbf{\Gamma}) &= (1/\sqrt{2})[G_{01}(t, \mathbf{\Gamma}) - \eta\alpha(\mathbf{\Gamma})G_{00}(t, \mathbf{\Gamma})], \\
K_{i0}(E, \mathbf{\Gamma}) &= -\sqrt{2}\eta t G_{01}(t, \mathbf{\Gamma}), \qquad (4.10) \\
K_{ii}(E, \mathbf{\Gamma}) &= G_{11}(t, \mathbf{\Gamma}) - \eta\alpha(\mathbf{\Gamma})G_{01}(t, \mathbf{\Gamma}), \\
K_{ij}(E, \mathbf{\Gamma}) &= G_{12}(t, \mathbf{\Gamma}) - \eta\alpha(\mathbf{\Gamma})G_{01}(t, \mathbf{\Gamma}), \quad i \neq j.
\end{aligned}$$

From (4.3), it follows that all the eigenvalues $k_i(E, \mathbf{\Gamma})$ of the matrix $K(E, \mathbf{\Gamma})$ which is specified by (4.10) are equal at each $t \neq 0$ when $\mathbf{\Gamma}$ is at H, their common value then being $1/4St$. Whence one concludes in the case of the present ferromagnets that when $\mathbf{\Gamma}$ coincides with this special point there is a quadruply degenerate bound state energy eigenvalue lying below $E_{min}(\mathbf{\Gamma})$, this eigenvalue being equal to $-J < 0$.

When $\mathbf{\Gamma} \in \Delta'$ and E lies either below $E_{min}(\mathbf{\Gamma})$ or above $E_{max}(\mathbf{\Gamma})$, three of the eigenvalues of this last matrix $K(E, \mathbf{\Gamma})$ are equal and the others are nondegenerate. Letting $\hat{k}_i(\mathbf{\Gamma}) \equiv k_i(E_{min}(\mathbf{\Gamma}), \mathbf{\Gamma})$ in the specific case when (4.10) holds, one finds over the latter range of $\mathbf{\Gamma}$ that

$$\hat{k}_1(\mathbf{\Gamma}) = \hat{k}_2(\mathbf{\Gamma}) = \hat{k}_3(\mathbf{\Gamma}) = \frac{I_0 + I_1 - 2I_2 - 1}{2S\alpha(\mathbf{\Gamma})},$$

$$\hat{k}_4(\mathbf{\Gamma}) = \frac{(I_0 - 1)}{4S}\left\{\frac{1}{\alpha(\mathbf{\Gamma})} - 2\eta + \left[\left(\frac{1}{\alpha(\mathbf{\Gamma})} - 2\eta\right)^2 + \frac{4\eta^2}{I_0 - 1}\right]^{1/2}\right\}. \qquad (4.11)$$

For the NN-coupled bcc ferromagnets being dealt with, the values of $\alpha(\mathbf{\Gamma})$ at which bound states occur at appropriate values of $E < E_{min}(\mathbf{\Gamma})$ when $\mathbf{\Gamma}$ roams over Δ' are now easily obtained from the inequalities $\hat{k}_1(\mathbf{\Gamma}) > 1$ and $\hat{k}_4(\mathbf{\Gamma}) > 1$. From (4.11) and the second of these inequalities, one finds for these ferromagnets at a given $S \geq \frac{1}{2}$ and $\mathbf{\Gamma} \in \Delta'$ that a nondegenerate bound state eigenvalue exists below $E_{min}(\mathbf{\Gamma})$ if and only if

$$\alpha(\mathbf{\Gamma}) < (I_0 - 1)/(2S + I_0 - 1) \qquad (4.12a)$$

at the S and $\mathbf{\Gamma}$ considered, and that the remaining type of bound state eigenvalues, the triply degenerate ones, exist in this energy range if and only if

$$\alpha(\mathbf{\Gamma}) < (I_0 + I_1 - 2I_2 - 1)/2S \qquad (4.12b)$$

at this S and $\mathbf{\Gamma}$.

Notice that the right-hand sides of (4.12a) and (4.12b) are positive, so that these bound state conditions are not empty, and that the first right-hand side is larger than the second. Hence the right-hand side of (4.12a) is the least upper bound of the values of $\alpha(\mathbf{\Gamma})$ over those vectors $\mathbf{\Gamma} \in \Delta'$ at which a bound state with an energy below $E_{min}(\mathbf{\Gamma})$ exists in the case (4.10) of present interest.[27] In other words, when $\mathbf{\Gamma}$ is confined to Δ' the first such states to emerge from the continuum in this case as $\alpha(\mathbf{\Gamma})$ decreases through this least upper bound (threshold value) are those pertaining to nondegenerate energy eigenvalues.

We now compare these exact results with rigorous trace-inequality bounds derived by the use of Theorems 3.3 and 3.5. In this comparison and henceforth, we shall employ the abbreviation $K(\mathbf{\Gamma}) \equiv K(E_{min}(\mathbf{\Gamma}), \mathbf{\Gamma})$.

In the case under discussion, we estimated the range of values assumed by $\alpha(\mathbf{\Gamma})$ over those $\mathbf{\Gamma} \in \Delta'$ at which no bound states with energies in the interval $E < E_{min}(\mathbf{\Gamma})$ occur by applying the conditions $\mathrm{Tr}K^2(\mathbf{\Gamma}) < 1$ and $\mathrm{Tr}K^4(\mathbf{\Gamma}) < 1$. When (4.10) obtains, these conditions hold at a given $\mathbf{\Gamma} \in \Delta'$ and a prescribed $S \geq \frac{1}{2}$ if and only if the respective inequalities

$$\alpha(\mathbf{\Gamma}) > \alpha_1(S), \qquad \alpha(\mathbf{\Gamma}) > \alpha_2(S),$$

do. At each such S, $\alpha_1(S)$ and $\alpha_2(S)$ are positive and clearly constitute upper bounds of the exact threshold value of $\alpha(\mathbf{\Gamma})$ furnished by (4.12a). It can be shown that these two bounds share an essential property of this exact value, namely, that they both tend to zero as $S \to \infty$.

We investigated the values assumed by $\alpha(\mathbf{\Gamma})$ at the vectors $\mathbf{\Gamma} \in \Delta'$ at which bound states with energies $E < E_{min}(\mathbf{\Gamma})$ occur for the ferromagnets under study by means of the conditions $\mathrm{Tr}K^4(\mathbf{\Gamma}) > \mathrm{Tr}K^3(\mathbf{\Gamma})$ and

TABLE II. Comparison of exact results with trace-inequality results for bound state thresholds, for bcc structures with NN interactions alone: Γ parallel to [100].

S	$\alpha_1(S)$	$\alpha_2(S)$	$\alpha_3(S)$	$\alpha_4(S)$	Right-hand Side of (4.12a)
$\frac{1}{2}$	0.4983	0.2944	0.2178	0.2598	0.2822
1	0.1969	0.1671	0.1364	0.1580	0.1643
$\frac{3}{2}$	0.1370	0.1175	0.0963	0.1120	0.1159

$\mathrm{Tr}K^5(\Gamma) > \mathrm{Tr}K^4(\Gamma)$. When (4.10) holds, these inequalities are satisfied at a specified Γ on this line and a chosen $S \geq \frac{1}{2}$ if and only if

$$\alpha(\Gamma) < \alpha_3(S), \quad \alpha(\Gamma) < \alpha_4(S),$$

respectively, at this Γ and S. At each S in this range, the quantities $\alpha_3(S)$ and $\alpha_4(S)$ are less than unity. Since it is obvious that $\alpha_3(S)$ and $\alpha_4(S)$ are lower bounds of the relevant exact threshold for every $S \geq \frac{1}{2}$, their positivity for each such S and the fact that they tend to zero as $S \to \infty$ follow trivially from previous remarks.

The values of the $\alpha_i(S)(i = 1, 2, 3, 4)$ are given in Table II for $S = \frac{1}{2}, 1, \frac{3}{2}$. It is seen from this table that $\alpha_2(S)$ is very close to the exact threshold result for these three spin values. This cannot be said for $\alpha_1(S)$ when $S = \frac{1}{2}$, but for $S = 1, \frac{3}{2}$, $\alpha_1(S)$ is reasonably close to the corresponding exact values, particularly in view of the fact that $\alpha_1(S)$ is obtained by using the lowest power of m stated in Theorem 3.3. Similarly, since $\alpha_3(S)$ results by using the lowest powers of p and q permitted by Theorem 3.5, we feel that $\alpha_3(S)$ is reasonably near to the pertinent exact values for $S = \frac{1}{2}, 1, \frac{3}{2}$. One also sees that $\alpha_4(S)$ is quite close to the corresponding exact value for these three choices of S.

We now return to our earlier results on the degeneracy of the eigenvalues of $K(E, \Gamma)$ in the case (4.10) at those pairs (E, Γ) such that Γ coincides with H or is a point of Δ', and which do not fulfill (2.32). These results are not explainable on the sole basis of the point group of Γ, which is the full cubic group O_h for the special points Γ and H and one of its subgroups, C_{4v}, for the special line Δ.[28] This is in sharp contrast with the corresponding situation for the NN-coupled sc structure, where, for pairs (E, Γ) such that Γ is a special point of the pertinent Brillouin zone or is a point of one of its special lines, and which do not obey (2.32), the degeneracies of the eigenvalues of this matrix can be accounted for by point-group arguments alone.

The spin-space groups discussed by Brinkman and Elliott[29] could well prove useful in ameliorating, or even in removing, the deficiencies of the point groups in these and other degeneracy problems encountered in the study of bound magnon states. Some symmetry properties of two-magnon states have been considered from this point of view,[30] but it is not known whether the problem of explaining bound state degeneracies is soluble by such means.

5. BCC STRUCTURE WITH NN INTERACTIONS: [111] AND [110] DIRECTIONS OF Γ

The methods of Sec. 3, whose effectiveness for these directions of Γ is illustrated in this section for the case of bcc ferromagnets with nonvanishing interactions between NN ions alone, have also been fruitfully applied to this case for the respective situations when Γ was on the special lines D and F on the boundary of Z_{bcc}. All further remarks on bound states and other matters in this section should be understood in the context of ferromagnets of this structure and mode of coupling.

For the respective cases when $\Gamma \in Z_{\mathrm{bcc}}$ was parallel to the [111] direction (line Λ and point P) and the [110] direction (line Σ), we computed the basic sets of Green's functions $G_{\mu\nu}(\hat{t}, \Gamma)$ mentioned in the Appendix by means of the techniques outlined therein. The elements of the pertinent matrix $K(E, \Gamma)$ were readily calculated at the desired values of E and Γ from a knowledge of the values of these basic Green's functions. The latter were evaluated to within an accuracy of at least 0.5×10^{-6} at the values of the parameter γ of Table I which were considered, except for values of γ pertaining to vectors Γ on Σ which were extremely close to the point $N(\pi - \gamma \lesssim 10^{-4}\pi)$, where the accuracy in question is judged to be about 10^{-5}. The adjacent values of the set of values of γ at which the above numerical integrations were performed for the [111] and [110] cases differed by amounts which varied from $2 \times 10^{-2}\pi$ to $10^{-4}\pi$. The differences between these adjacent eigenvalues were carefully chosen small enough to enable us, in particular, to safely reach the conclusions of this section relative to the respective ranges of validity of various trace inequalities employed to investigate the existence of bound states at energies $E < E_{min}(\Gamma)$ in the stated ranges of Γ, and to reliably arrive at the exact bound state result (5.3). The negative conclusion mentioned below concerning bound states with energies $E > E_{max}(\Gamma)$ at the specified Γ vectors was independent of any such choice.

Purely for the sake of completeness, we mention that only qualitative arguments are required to establish that, for the ferromagnets being investigated, bound states exist for each $S \geq \frac{1}{2}$ at some $E < E_{min}(\Gamma)$ for points Γ of Σ, D, and F which are sufficiently close to the point N on the zone boundary. (Arguments of the same type are effective in establishing analogous qualitative results for the [100] case dealt with in Sec. 4 and also when Γ is any point of line G or coincides with point N, but they are of little interest in these three cases, since then the two-magnon bound state problem can be solved exactly in simple closed form[31]). Indeed, if Γ is a special point of Z_{bcc} different from H or is a point of any special line of this zone, it can be proved with the aid of (2.38), (4.3), and Table I that $K(\Gamma)$ has the (finite) positive eigenvalue

$$k(\Gamma) = (1/2S)[G_{11}(\hat{t}, \Gamma) - G_{12}(\hat{t}, \Gamma)]$$

for every $S \geq \frac{1}{2}$ when only NN interactions are operative. Employing (2.28), (2.36), (4.2), (4.3), and Table I, one concludes that $k(\Gamma) \to \infty$ as Γ tends to point N along Σ or D or to point H along F. From this and the monotonicity and continuity properties of the $k_\mu(E, \Gamma)$, the above statement relative to the existence of bound states for suitable points Γ of Σ, D, and F follows.

We now turn to the study of the bound state problem of interest for the points Γ of Λ', which we define as the line consisting of the points of Λ and of the point P. We shall then deal with the corresponding problem for line Σ.

Line Λ'

In the case of the ferromagnets under discussion, we shall prove that whenever Γ is on this line no bound states exist for any $S \geq \frac{1}{2}$ when $E > E_{max}(\Gamma)$, and shall summarize the numerical trace-inequality results which exclude bound states from the range $E < E_{min}(\Gamma)$

at all such Γ for a set of values of S which presumably include all those of practical interest.

For these ferromagnets, a sufficient condition for the nonexistence of a bound state at an energy above $E_{max}(\Gamma)$ for a given $S \geq \frac{1}{2}$ and a prescribed $\Gamma \in \Lambda'$ is that the inequality

$$G_{00}(\hat{t},\Gamma) \leq 2S \tag{5.1}$$

be satisfied at the S and Γ, as follows from considerations similar to ones used to establish the analogous condition (4.9). We proceed to show that (5.1) obtains at the desired parameter values by means of an argument whose only appeal to numerical work is in the evaluation of $G_{00}(\hat{t},\Gamma)$ at a single value of Γ. It can be shown purely analytically that $G_{00}(\hat{t},\Gamma)$ is bounded at all $\Gamma \in \Lambda'$ and that it attains its maximum over all such Γ at the unique value $\Gamma = \Gamma_c \equiv (\gamma_c,\gamma_c,\gamma_c), \gamma_c$ being the only root of

$$\cos(\gamma/4) = \sqrt{2/3}$$

in the interval $0 < \gamma < \pi$. Numerical computation yields:

$$G_{00}(\hat{t},\Gamma) = 0.950819 \text{ at } \Gamma = \Gamma_c.$$

Therefore (5.1) holds for all $S \geq \frac{1}{2}$ when $\Gamma \in \Lambda'$.

Preliminary to discussing the existence of bound states at energies $E < E_{min}(\Gamma)$ for vectors Γ of this last type, we mention a phenomenon pertaining to the unique nonpositive eigenvalue $k_0(\Gamma)$ of $K(\Gamma) \equiv K(E_{min}(\Gamma), \Gamma)$. Our trace calculations revealed that $|k_0(\Gamma)| > 1$ when $S = \frac{1}{2}$, in particular when Γ was located in certain segments of Λ' and Σ, these segments being relatively sizeable. Hence Theorem 3.3 proves unsatisfactory for the investigation of the regions of nonexistence of bound states at energies $E < E_{min}(\Gamma)$ in such cases. On the other hand, Theorem 3.4 was found to be a very efficient tool in excluding these states at such energy values over substantial ranges of Γ in the examples alluded to below, even when $S = \frac{1}{2}$. When applied to the present type of ferromagnets, Theorem 3.3 yielded good results in the cases $S > \frac{1}{2}$ which were examined.

Our investigation of the existence of bound states at energies below $E_{min}(\Gamma)$ for $\Gamma \in \Lambda'$, based on the last two mentioned theorems, led to the following conclusions. In the case $S = \frac{1}{2}$, the inequality

$$\lambda_{pq}(\Gamma) \leq 1$$

held for p and q as small as 3 and 4, respectively, over the latter range of Γ, where $\lambda_{pq}(\Gamma) \equiv \lambda_{pq}(E_{min}(\Gamma), \Gamma)$ [$\lambda_{pq}(E,\Gamma)$ was defined in (3.3a)]. The inequality

$$\mathrm{Tr}K^2(\Gamma) < 1$$

was amply satisfied at all the Γ just alluded to when $1 \leq S \leq 5$. We therefore conclude that the ferromagnets we are now dealing with have no bound states at any such Γ in the energy range $E < E_{min}(\Gamma)$ when $\frac{1}{2} \leq S \leq 5$.

For $\Gamma = (\gamma,\gamma,\gamma) \in \Lambda'$, the $G_{\mu\nu}(\hat{t},\Gamma)$ exhibit a sharp upward cusp as functions of the real variable γ, the tip of this cusp being located at $\gamma = \gamma_c$. This behavior is responsible for the fact that, in particular at each S in the last cited range, the largest eigenvalue of $K(\Gamma)$ for the latter ferromagnets attains a sharp maximum at Γ_c as Γ varies along Λ'. When $S = \frac{1}{2}$, this maximum value is not too far from unity, being then equal to 0.8318. Hence it is natural to conjecture that these ferromagnets possess a two-magnon resonance at a (complex) value of E "near" $E_{min}(\Gamma_c)$ for the latter spin value. The truth or falsity of this conjecture remains to be determined.

Line Σ

The use of Theorems 3.3 and 3.4 eliminated the possibility of bound states with energies $E < E_{min}(\Gamma)$ for $\Gamma \in \Sigma$, except at values of Γ on this line extremely close to the zone boundary. For example, the application of the inequality

$$\lambda_{pq}(\Gamma) \leq 2$$

with $p = 5, q = 4$ for $S = \frac{1}{2}$, of this same inequality with $p = 3, q = 2$ for $S = 1$, and even of the inequality (3.4) with $m = 1$ for $\frac{1}{2} \leq S \leq 5$, showed that no bound states were possible at these energies for vectors $\Gamma = (\gamma,\gamma,0)$ such that

$$\begin{aligned} &0 < \gamma \leq 0.9880\pi, \quad &&\text{if } S = \tfrac{1}{2}, \\ &0 < \gamma \leq 0.9999\pi, \quad &&\text{if } 1 \leq S \leq 5. \end{aligned} \tag{5.2}$$

Results almost as good as the respective ones in (5.2) were obtained for the values of Γ on Σ at which no bound states exist by applying (3.5a) with $p = 3, q = 4$ for $S = \frac{1}{2}$, (3.5b) with $p = 3, q = 2$ at this same spin value, and (3.4) with $m = 1$ for $S = 1$. Incidentally, the upper bound on γ in the second line of (5.2) is the highest value of this parameter at which the Green's function computations of interest in this subsection were carried out. The numerical calculation of the $G_{\mu\nu}(\hat{t},\Gamma)$ for line Σ by the approach of the Appendix becomes increasingly tedious beyond this value.

For $S = \frac{1}{2}$, we calculated the eigenvalues of $K(\Gamma)$ directly on the small portion of Σ where trace-inequality techniques failed. At each $S \geq \frac{1}{2}$, one can show that a bound state with energy in the range $E < E_{min}(\Gamma)$ exists at a vector $\Gamma = (\gamma,\gamma,0)$ on Σ if and only if γ lies in a certain open interval which, to the present accuracy, is

$$0.9958\pi < \gamma < \pi, \quad \text{if } S = \tfrac{1}{2}, \tag{5.3}$$

for the ferromagnets investigated in this section. Comparing the respective stated results expressed by (5.3) and by the first line of (5.2), one sees that the latter trace-inequality result is in excellent agreement with the corresponding exact one.

For these ferromagnets, it was not possible to exclude bound states at energies above $E_{max}(\Gamma)$ for all $\Gamma \in \Sigma$ by the theorems of Sec. 3. Since a direct eigenvalue approach to this question would have been rather tedious, the matter was not pursued any further.

ACKNOWLEDGMENTS

We wish to thank Mr. S. Podgor for programming the computation of the required quantities and Mr. J. B. Langworthy for conversations on pertinent numerical matters. A discussion with Professor J. Callaway is gratefully acknowledged.

APPENDIX. NUMERICAL EVALUATION OF GREEN'S FUNCTIONS FOR THE BCC STRUCTURE WITH NN INTERACTIONS: [111] AND [110] DIRECTIONS OF Γ

Before discussing in more detail the procedures which we employed to reduce to convenient numerical proportions the problem of computing the functions $G_{\mu\nu}(t,\Gamma)$ in (2.34) for the case $t = \hat{t}(\Gamma)$ when Γ had one

of these respective directions, we shall summarize these procedures schematically. The $G_{\mu\nu}(\hat{t},\Gamma)$ with $\mu,\nu = 0,1,2$, or extensions of these functions, were expressed as linear combinations of single integrals, whose integrands were known exactly in closed form in terms of complete elliptic integrals. After appropriate transformations to remove the infinite singularities in these integrands, the single integrals in question were computed by Gaussian quadratures on a CDC 3800 computer at the Naval Research Laboratory. From the values of these $G_{\mu\nu}(\hat{t},\Gamma)$ or of the alluded to extensions the remaining $G_{\mu\nu}(\hat{t},\Gamma)$ were computed for vectors $\Gamma \in Z_{\rm bcc}$ in the [111] and [110] directions via simple algebraic formulas, exception made of the case when Γ had the value implicitly specified below. Similar procedures are applicable to the calculation of the $G_{\mu\nu}(\hat{t},\Gamma)$ when Γ is on D or F.

By first integrating over the variable ρ_3 in (2.34) and then making the change of variables

$$x = \tfrac{1}{2}(\rho_1 + \rho_2), \quad y = \tfrac{1}{2}(\rho_1 - \rho_2),$$

one finds in the present case of bcc structures with NN couplings only that

$$G_{\mu\nu}(t,\Gamma) = \frac{2}{\pi^2}\int_0^\pi\int_0^\pi \frac{\cos\rho_\mu(x,y)\,\cos\rho_\nu(x,y)\,dx\,dy}{[(t - a\cos x\cos y)^2 - (b + c\cos^2 x)]^{1/2}}, \quad \text{if } \mu,\nu = 0,1,2 \text{ and } |t| \le \hat{t}(\Gamma), \tag{A1}$$

whenever Γ lies on any of the special points or lines of $Z_{\rm bcc}$ at which the finiteness condition of Sec. 3 holds. Here a, b, and c depend on Γ but not on x, y, μ, or ν. Until further notice, the following remarks refer to vectors Γ fulfilling all the conditions just stated and to values of t as specified in (A1).

The $G_{\mu\nu}(t,\Gamma)$ with $\mu,\nu = 0,1,2$ are expressible as linear combinations of double integrals I_{rs} of type (A1), but with $\cos\rho_\mu\,\cos\rho_\nu$ replaced by $\cos^r x\,\cos^s y$, where r and s are integers such that $0 \le r, s \le 2$. Effecting either the x or y integration in a given I_{rs} whose subscripts obey these requirements, the double integral in question reduces to the form[32]

$$\int_0^{\pi/2}[l(z)K(k(z)) + m(z)E(k(z)) + n(z)\Pi(\alpha^2(z),k(z))]dz, \tag{A2}$$

which, for $r = s = 0$, collapses to the simpler representation

$$G_{00}(t,\Gamma) = \int_0^{\pi/2} f(z)K(k(z))dz \tag{A2}$$

for the indicated Green's function given by (2.34), where z denotes the variable x or y not integrated over, and where the dependence on various parameters and indices has been suppressed. The functions $l(z), m(z), n(z)$, and $f(z)$, as well as the moduli $k(z)$ and parameters $\alpha^2(z)$ of the complete elliptic integrals $K(k(z))$, $E(k(z))$, and $\Pi(\alpha^2(z),k(z))$ of the first, second, and third kinds, respectively, are elementary functions which involve t and Γ, and whose functional forms depend on whether the x or y integration was performed to arrive at (A2) or (A2′). However, $k(z)$ and $\alpha^2(z)$ are independent of r and s.

Henceforth we shall limit ourselves to the situation when $t = \hat{t}(\Gamma)$. In this situation, the denominator of (A1) has zeros in the square $0 \le x, y \le \pi$ of integration. Because of this vanishing, the integrand of any given integral of type (A2) corresponding to an I_{rs} with indices satisfying the conditions stated in the last paragraph has an infinite singularity at $z = 0$ (at a value of z in the interval $0 \le z \le \pi/4$) in the case $z = x (z = y)$. This singularity is of logarithmic type for either $z = x$ or $z = y$ if Γ satisfies the requirements mentioned in (A1) and, in addition, is different from the vector Γ_c defined in Sec. 5. Elsewhere in the interval $0 \le z \le \pi/2$, each of the latter integrands is infinitely differentiable in z for both of these choices of this variable.

In the remainder of this Appendix, we shall restrict our attention first to the case of the [111] direction and then to that of the [110] direction.

Line Λ'

Whenever Γ is a point of Λ' or coincides with the point Γ at the center of $Z_{\rm bcc}$, we set $G_{\mu\nu}(t,\Gamma) \equiv G_{\mu\nu}(\beta)$, where β is the value of the parameter α_1 in Table I at the value of Γ of interest. Hence $1/\sqrt{2} \le \beta \le 1$ for this range of Γ. The point $\Gamma(\beta - 1)$ was considered in the present work in order to check the values of $G_{\mu\nu}(1)$ obtained by the method which will now be explained with those obtained from the exact results (4.8).

Equations (2.38) and elementary symmetry arguments entail that all the $G_{\mu\nu}(\beta)$, are determined by linear combinations of $G_{00}(\beta), G_{01}(\beta), G_{11}(\beta)$, and $G_{12}(\beta)$ at any $1/\sqrt{2} \le \beta \le 1$ different from $\sqrt{3}/2$. Certain of these combinations are not numerically stable for β close enough to $\sqrt{3}/2$, but this can be remedied by interpolatory procedures to an accuracy sufficient for our purposes.

In the interval

$$\sqrt{2/3} \le \beta \le 1, \tag{A3a}$$

the four basic $G_{\mu\nu}(\beta)$ just mentioned were calculated by always making the choice $z = x$, to avoid the very awkward formulas obtained for the latter $G_{\mu\nu}(\beta)$ in the subinterval $\sqrt{2/3} \le \gamma \le \sqrt{3}/2$ of the interval (A3a) when $z = y$.

To compute the function $G_{00}(\beta)$ accurately in the latter interval, a modification of the representation (A2′) was employed, namely, a suitable term was subtracted from the pertinent integrand to eliminate the logarithmic singularity occurring in the latter and a compensating term was added, whose integral over the relevant interval was known exactly. The regularized integral was then evaluated numerically. In this and the remaining computations below, the pertinent elliptic integrals were calculated by a Landen-transformation subroutine in the case of those of the first kind and by Bartky-transformation subroutines for those of the second and third kinds.

For $(\mu,\nu) = (0,1), (1,1), (1,2)$, we wrote

$$G_{\mu\nu}(\beta) = G_{00}(\beta) + H_{\mu\nu}(\beta). \tag{A4}$$

Because the above infinite singularities occur only at the value $z = x = 0$ of the chosen integration variable when (A3a) holds, we were able to represent these $H_{\mu\nu}(\beta)$ in this range of β as sums of single integrals of type (A2) with bounded integrands. For each such β, outside of having the stated differentiability properties for $0 < x \le \pi/2$, these integrands are asymptotically proportional to $x^2\log(1/x)$ as $x \to +0$. The $H_{\mu\nu}(\beta)$ in question were easily evaluated to the desired accuracy.

Another method was devised in the interval

$$1/\sqrt{2} \le \beta \le \sqrt{2/3}. \tag{A3b}$$

As an added check of our numerical work, this interval was purposely selected to overlap the interval (A3a), in order to compare the values of the $G_{\mu\nu}(\beta)$ at $\beta = \sqrt{2/3}$, i.e., at $\Gamma = \Gamma_c$, obtained by the previously outlined method and the one summarized below.

In the interval (A3b), we reduced $G_{00}(\beta), G_{01}(\beta), G_{11}(\beta)$, and $G_{12}(\beta)$ directly to the form (A2), always with $z = y$ and without using decompositions of the type (A4). If we had performed the same reduction for any of these $G_{\mu\nu}(\beta)$ at any given β in this interval, but with $z = x$, an infinite singularity of the resulting integrand would have occurred at a point of the domain of integration which would have varied with the value of β selected, rather than occurring at the numerically more advantageous fixed value $y = 0$, as it does when the choice $z = y$ is made.

In the case $z = y$, the integrands of the representations of type (A2) of $G_{00}(\beta), G_{01}(\beta), G_{11}(\beta)$, and $G_{12}(\beta)$ are asymptotically proportional to $\log(l/y)$ as $y \to +0$, if β is any given number such that $1/\sqrt{2} \leq \beta < \sqrt{2/3}$ and to $1/\sqrt{y}$ if $\beta = \sqrt{2/3}$. The cusp possessed by each of the $G_{\mu\nu}(\beta)$ at $\beta = \sqrt{2/3}$, a fact already mentioned in Sec. 5, is connected with this asymptotic behavior in the interval (A3b), as well as with asymptotic phenomena of each of the pertinent integrands taking place within the remainder of Λ'. Both of the types of infinite singularities mentioned in the penultimate sentence can be removed by a change

$$y = w^{2m}, \quad m = 1, 2, \cdots, \tag{A5}$$

of the integration variable. For each positive integer m, this change of variable allows us to express $G_{00}(\beta)$, $G_{01}(\beta), G_{11}(\beta)$, and $G_{12}(\beta)$ as one dimensional integrals with integrands possessing the following properties. They are infinitely differentiable in w over the transformed interval $0 \leq w \leq (\pi/2)^{1/m}$ of integration when $\gamma = \sqrt{2/3}$, and when $1/\sqrt{2} \leq \gamma < \sqrt{2/3}$ they have this property at all $w \neq 0$ in this interval and are asymptotically proportional to $w^{2m-1} \log(1/w)$ as $w \to +0$.

We evaluated the latter four Green's functions numerically in a very satisfactory manner in the interval (A3b) by the transformation (A5) with $m = 2$. When using this approach, it was necessary to use appropriate, easily derivable, formulas for calculating the moduli and parameters of the relevant elliptic integrals at the smaller values of w required in the Gauss integration method, in order to avoid numerical instability difficulties.

Line Σ

If Γ is a point of Σ or is coincident with point Γ, we write $G_{\mu\nu}(\vec{l}, \Gamma) \equiv \tilde{G}_{\mu\nu}(\gamma), \gamma$ being the value of the parameter α_1 of Table I at the Γ in question. Therefore $0 < \gamma \leq 1$ when Γ runs over the indicated range. Point $\Gamma(\gamma = 1)$ was considered in the calculations outlined below for a reason parallel to one stated in the previous subsection. For reasons which will become clear soon, it is convenient to extend the domain of definition of the $\tilde{G}_{\mu\nu}(\gamma)$ $(\mu, \nu = 0, 1, 2)$ to $\gamma > 0$, e.g., by means of the representation (A1). All of the previously stated results applying on Σ hold equally well at any $\gamma > 0$.

By arguments similar to those alluded to in connection with Λ', the values of all the $\tilde{G}_{\mu\nu}(\gamma)$ at each $0 < \gamma < 1$ are calculable as linear combinations of $\tilde{G}_{00}(\gamma), \tilde{G}_{01}(\gamma), \tilde{G}_{11}(\gamma), \tilde{G}_{12}(\gamma), \tilde{G}_{33}(\gamma)$, or $\tilde{G}_{00}(\gamma), \tilde{G}_{03}(\gamma), \tilde{G}_{11}(\gamma), \tilde{G}_{33}(\gamma), \tilde{G}_{34}(\gamma)$, at the γ in question. However, some of the linear combinations giving the $\tilde{G}_{\mu\nu}(\gamma)$ in terms of the first of these sets of Green's functions are numerically unstable at small enough values of γ. This is disagreeable, since bound states occur on Σ only for extremely small values of γ. However, the $\tilde{G}_{\mu\nu}(\gamma)$ can be expressed in a stable way in terms of the second set of Green's functions for all Γ on Σ. We thus determined numerically the five functions of this latter set in the range $0 < \gamma \leq 1$. To do this, it was only necessary to calculate $\tilde{G}_{00}(\gamma), \tilde{G}_{01}(\gamma), \tilde{G}_{11}(\gamma)$, and $\tilde{G}_{12}(\gamma)$ at suitable values of $\gamma > 0$ and then to use the easily provable formulas

$$\tilde{G}_{03}(\gamma) = (1/\gamma)\tilde{G}_{01}(1/\gamma),$$

$$\tilde{G}_{33}(\gamma) = (1/\gamma)\tilde{G}_{11}(1/\gamma),$$

$$\tilde{G}_{34}(\gamma) = (1/\gamma)\tilde{G}_{12}(1/\gamma),$$

which obtain at each $\gamma > 0$.

In the latter range of γ, the choices $z = x$ and $z = y$ are about equally convenient from a numerical viewpoint.

The required $\tilde{G}_{\mu\nu}(\gamma)$ were readily computed to the desired precision by methods parallel to those employed in the case (A3a).

*Preliminary results of the questions dealt with in this paper were reported by A. W. Sáenz and W. W. Zachary, Bull. Am. Phys. Soc. **13**, 92 (1968) and A. W. Sáenz, Symposium on Current Problems in Neutron Scattering, Rome, Italy, September 24–27, 1968 (unpublished).

[1] A. W. Sáenz, W. W. Zachary, and S. Podgor, Bull. Am. Phys. Soc. **15**, 1368 (1970).

[2] In the case of magnon bound states (resonances), the adjective "low lying" indicates that the corresponding energy eigenvalue (real part of the energy) differs from the lower edge of the pertinent free magnon continuum by an energy which is small in comparison with a typical value of the width of the continuum.

[3] In regard to the relevance of the nonexistence of low-lying magnon bound states to the validity of low-temperature spin wave calculations, see, e.g., M. Wortis, Phys. Rev. **138**, A1126 (1965) and T. Morita and T. Tanaka, J. Math. Phys. **6**, 1152 (1965).

[4] R. Silberglitt and A. B. Harris, Phys. Rev. Lett. **19**, 30 (1967); Phys. Rev. **174**, 640 (1968).

[5] J. B. Torrance, Jr. and M. Tinkham, Phys. Rev. **187**, 587 (1969).

[6] M. F. Thorpe, Phys. Rev. B **4**, 1608 (1971).

[7] F. J. Dyson, Phys. Rev. **102**, 1217 (1956).

[8] J. Hanus, Phys. Rev. Lett. **11**, 336 (1963).

[9] M. Wortis, Phys. Rev. **132**, 85 (1963); N. Fukuda and M. Wortis, J. Phys. Chem. Solids **24**, 1675 (1963).

[10] In the present paper, we have found it more convenient to refer to specific points, lines, and directions of symmetry than to the corresponding equivalence classes. Naturally, all statements about bound two-magnon states and allied topics made here for the case when Γ lies along some such point, line, or direction hold at the vectors Γ' related to Γ by the pertinent symmetry group operations.

[11] R. G. Boyd, Ph.D. Thesis, Univ. of California, Riverside, 1965; R. G. Boyd and J. Callaway, Phys. Rev. **138**, A1621 (1965).

[12] W. M. Shaw, Bull. Am. Phys. Soc. **13**, 501 (1968).

[13] The beginnings of a theory of three interacting magnons has been given by C. K. Majumdar, Phys. Rev. B **1**, 287 (1970), and C. K. Majumdar and G. Mukhopadyay, Phys. Lett. A **31**, 321 (1970).

[14] For the sc structure, our definition of $Z_N(\Gamma)(\Gamma \in Z_N)$ agrees with that of the corresponding "reduced zone" F of Wortis, Ref. 9.

[15] Exception made of one-dimensional ferromagnets [See Wortis, Ref. 9, Appendix B], only arguments of a heuristic nature appear to have been given in the literature relative to the spacing of consecutive two-magnon eigenvalues of $H^{(N)}$ of given Γ within the corresponding band (2.15) for large N.

[16] W. Ledermann, Proc. R. Soc. A **182**, 362 (1944).

[17] H. Rollnik, Z. Phys. **145**, 639 (1956).

[18] S. Weinberg,,Phys. Rev. **131**, 440 (1963). See also R. G. Newton, *Scattering Theory of Waves and Particles* (McGraw-Hill, New York, 1966), Chap. 9.

[19] The stated differentiability properties of the $k_\mu(E, \Gamma)$ which were

needed in our proof of (2.40), as well as the existence of eigenfunctions of $K(E, \Gamma)$ with differentiability attributes strong enough to make meaningful the operations involved in that proof, were derived by combining the relevant facts in pp. 63–65 of T. Kato, *Perturbation Theory for Linear Operators* (Springer, New York, 1966) with straightforward arguments.

[20]One of us (A.W.S.) intends to present this proof in a forthcoming Naval Research Laboratory Report.

[21]J. Schwinger, Proc. Natl. Acad. Sci. USA **47**, 122 (1961). See also Weinberg, Ref. 18, and G. C. Ghirardi and A. Rimini, J. Math. Phys. **6**, 40 (1965).

[22]See, for example, G. H. Hardy, J. E. Littlewood, and G. Pólya, *Inequalities* (Cambridge U. P., Cambridge, 1958), 2nd ed., p. 28, Theorem 19.

[23]See Theorem 18.5, p. 27 of C. C. MacDuffee, *The Theory of Matrices* (Chelsea, New York, 1946).

[24]G. N. Watson, Q. J. Math. **10**, 266 (1939).

[25]G. S. Joyce, J. Math. Phys. **12**, 1390 (1970).

[26]Wortis, Ref. 9, proved a nonexistence result, similar to this one of ours, for an analogous class of isotropic sc ferromagnets, for the [111] direction of Γ. His result can be established very simply by the approach used to prove ours in the text.

[27]Compare with the parallel results of Wortis, Ref. 9, for the analogous sc case alluded to in the preceding footnote.

[28]L. P. Bouchaert, R. Smoluchowski, and E. P. Wigner, Phys. Rev. **50**, 58 (1936). In fact, the maximum dimensionality of an irreducible representation of C_{4v} is two and that of an irreducible representation of O_h is three.

[29]W. F. Brinkman and R. J. Elliott, Proc. R. Soc. A **294**, 343 (1966); W. F. Brinkman, J. Appl. Phys. **38**, 939 (1967).

[30]P. A. Fleury and R. Loudon, Phys. Rev. **166**, 514 (1968). See also R. J. Elliott, M. F. Thorpe, G. F. Imbusch, R. Loudon, and J. B. Parkinson, Phys. Rev. Lett. **21**, 147 (1968) and references contained therein. More recent references can be obtained from P. A. Fleury and H. J. Guggenheim, Phys. Rev. Lett. **24**, 1346 (1970).

[31]For the NN-coupled bcc structures under consideration, the $G_{\mu\nu}(t, \Gamma)$ are expressable as linear combinations of complete elliptic integrals when Γ is on G or at N, in parallel with what was found by Wortis, Ref. 9, for analogous sc ferromagnets when Γ was on the pertinent zone boundary.

[32]One can reduce the I_{rs} to the form (A2) by employing a new variable of integration $u = \cos x$ or $v = \cos y$ and using integration formulas given, e.g., by P. F. Byrd and M. D. Friedman, *Handbook of Elliptic Integrals for Engineers and Physicists* (Springer-Verlag, Berlin, 1954).

Deformations of inhomogeneous classical Lie algebras to the algebras of the linear groups

Charles P. Boyer and Kurt Bernardo Wolf

C. I. M. A. S. Universidad Nacional Autónoma de México, México 20, D. F., México
(Received 6 March 1973)

We study a new class of deformations of algebra representations, namely, $i_2so(n) \Rightarrow sl(n, \mathbb{R})$, $i_2u(n) \Rightarrow sl(n, \mathbb{C}) \oplus u(1)$ and $i_2sp(n) \oplus sp(1) \Rightarrow sl(n, Q) \oplus sp(1)$. The new generators are built as commutators between the Casimir invariant of the maximal compact subalgebra and a second-rank mixed tensor. These algebra deformations are related to multiplier representations and manifold mappings of the corresponding Lie groups. Behavior of the representations under Inönü–Wigner contractions is exhibited. Through the use of these methods we can construct a principal degenerate series of representations of the linear groups and their algebras.

1. INTRODUCTION

The closely related concepts of expansion and deformation of Lie algebras has been developed in both the physics and mathematics literature. In physics, expansions first appeared as a way of building symplectic algebras $sp(n, \mathbb{R})$ from the position and momentum operators with the canonical commutation relations,[1] and later by deforming the Poincaré algebra to the de Sitter algebra[2] as well as obtaining possible dynamical algebras for various physical systems.[3,4] Indeed, these latter types of deformations have been performed for inhomogeneous orthogonal,[2,5,6] unitary[5,6] and symplectic[6,7] Lie algebras using a specific type of deformation, i.e., $iso(n) \Longrightarrow so(n, 1)$, $iu(n) \oplus u(1) \Longrightarrow u(n, 1)$, $isp(n) \oplus sp(1) \Longrightarrow sp(n, 1)$ and other noncompact forms. These deformations have then been applied to various problems in representation theory[5,8] and shown by Gilmore[9] to constitute a well-defined family of deformations in which the coset space of the deformed algebra in the Cartan decomposition is of rank one.

In this article we present a family of deformations of representations of Lie algebras on homogeneous spaces of rank one (spheres), but where the rank of the coset spaces of the deformed algebras in the Cartan decomposition is higher. Specifically, in Sec. 2 we treat the cases $i_2so(n) \Longrightarrow sl(n, \mathbb{R})$, $i_2u(n) \Longrightarrow sl(n, \mathbb{C}) \oplus u(1)$ and $i_2sp(n) \oplus sp(1) \Longrightarrow sl(n, \mathbb{Q}) \oplus sp(1)$. We develop these cases separately so as to provide a clearer formulation for the reader who is not so familiar with the complications of the quaternionic field, which would be necessary in a general formulation. We then discuss in Sec. 3 the connection between the deformations of these algebras and the existence of corresponding multiplier representations[10,11] of the groups $SL(n, \mathbb{R})$, $SL(n, \mathbb{C}) \otimes U(1)$ and $SL(n, \mathbb{Q}) \otimes Sp(1)$ on the homogeneous spaces corresponding to the real, complex and quaternionic spheres. In Sec. 4 we show that the Inönü–Wigner contraction[12] of the representations of these groups with respect to the maximal compact subgroups are the groups $I_2SO(n)$, $I_2SU(n) \otimes U(1)$, and $I_2Sp(n) \otimes Sp(1)$.

2. DEFORMATIONS OF INHOMOGENEIZATIONS OF THE CLASSICAL LIE ALGEBRAS

Since we will be interested in deformations[12] and expansions[9] which are representation-dependent, we begin with suitable definitions of expansions and deformations of representations in which nothing is said about the abstract Lie algebra. Let ϕ be a representation of a Lie algebra $\mathfrak{A}$, i.e., a homomorphism of $\mathfrak{A}$ into some suitable defined vector space, which for our purposes can be taken as the space of infinitely differentiable functions over spheres. An *expansion* of the representation ϕ is a mapping $\phi \to \psi_\lambda$ such that the ψ_λ's form a representation of a Lie algebra $\mathfrak{A}'$. Moreover, if the condition $\psi_\lambda \xrightarrow{\lambda \to 0} \phi$ is satisfied, the expansion is said to be a *deformation*. The deformation of an inhomogeneous algebra can be thought of as the inverse of contraction.[12] It is seen that the requirement that the ψ_λ's form a Lie algebra places severe restrictions on the possible mappings ψ_λ. Such restrictions have an elegant formulation in terms of Lie algebra cohomology theory,[13] however, rather than attempt the general formulation here, we will discuss a family of specific examples of representation-dependent deformations of inhomogeneizations of the classical Lie algebras.

A. $i_2so(n) \Rightarrow sl(n, \mathbb{R})$

Consider the Lie algebra $so(n)$ of the orthogonal group whose generators satisfy the well-known commutation relations[14]

$$[M_{\mu\nu}, M_{\rho\sigma}] = \delta_{\nu\rho}M_{\mu\sigma} - \delta_{\mu\rho}M_{\nu\sigma} - \delta_{\nu\sigma}M_{\mu\rho} + \delta_{\mu\sigma}M_{\nu\rho}, \quad (2.1)$$

which preserve the usual metric in real n-space $\mathbb{R}^n$, so that the Greek indices take values $1, \ldots, n$. We adjoin now to this algebra a set of commuting n-dimensional second-rank symmetric tensors $P_{\mu\nu} = P_{\nu\mu}$. We thus arrive at a Lie algebra which we denote by $i_2so(n)$, which is characterized, along with Eq. (2.1) by

$$[M_{\mu\nu}, P_{\rho\sigma}] = \delta_{\nu\rho}P_{\mu\sigma} - \delta_{\mu\rho}P_{\nu\sigma} + \delta_{\nu\sigma}P_{\mu\rho} - \delta_{\mu\sigma}P_{\nu\rho}, \quad (2.2a)$$

$$[P_{\mu\nu}, P_{\rho\sigma}] = 0. \quad (2.2b)$$

The set of $\frac{1}{2}n(n+1)$ generators P constitute the maximal Abelian ideal of $i_2so(n)$.

The technique for deformation now consists of taking the commutator of the Casimir operator Φ of the original algebra $so(n)$ with the P's. Specifically, we consider the following members of the enveloping algebra of $i_2so(n)$:

$$N_{\mu\nu} \equiv \tfrac{1}{2}[\Phi, P_{\mu\nu}] + \tau P_{\mu\nu}, \quad (2.3)$$

where $\Phi \equiv -\frac{1}{2}M_{\mu\nu}M_{\nu\mu}$ and τ is an arbitrary complex number.

As Φ commutes with all of $so(n)$, it follows that the N's transform under $so(n)$ as the P's, i.e., they satisfy Eq. (2.2a) with $P_{\mu\nu}$ replaced by $N_{\mu\nu}$. However, if we consider the analog of Eq. (3.2b), that is, the commutator of two N's, we find that in general, (for any choice of τ other than the contraction limit $\tau \to \infty$), the N's do not close into a finite-dimensional Lie algebra. This is to

be contrasted with the better-known expansion[2,9] $iso(n) \Longrightarrow so(n,1)$ where the algebra closes modulo a normalization factor. We can, however, obtain a representation of a Lie algebra if we impose some further restrictions. We choose the following representation[15] for $P_{\mu\nu}$ and $M_{\mu\nu}$: $P_{\mu\nu} = x_\mu x_\nu/x^2$ where the x_μ's commute, $x^2 = x_\mu x_\mu$ and $M_{\mu\nu} = x_\mu\partial_\nu - x_\nu\partial_\mu + \sigma_{\mu\nu}$ with $[\sigma_{\mu\nu}, x_\lambda] = 0$, where we have introduced $\partial_\mu \equiv \partial/\partial x_\mu$. It is then found that the algebra will close if and only if $\sigma_{\mu\nu}$ vanishes. This means that, in contrast with the $iso(n) \Longrightarrow so(n,1)$ expansion, we do not have the freedom to construct an additional vector space structure upon the representations described by Eq. (2.3), i.e., vector bundles over the sphere. This limits the possible representations one can construct to degenerate ones.[16] In a straightforward manner one can then verify that

$$[N_{\mu\nu}, N_{\rho\sigma}] = \delta_{\nu\rho}M_{\mu\sigma} + \delta_{\mu\rho}M_{\nu\sigma} + \delta_{\nu\sigma}M_{\mu\rho} + \delta_{\mu\sigma}M_{\nu\rho}. \quad (2.4)$$

In order to see more clearly the structure of the algebra spanned by the M's and N's it is convenient to construct the traceless operators

$$X_{\mu\nu} \equiv \frac{1}{2}(M_{\mu\nu} + N_{\mu\nu}) - \frac{1}{2n}\delta_{\mu\nu}\,\mathrm{Tr}N$$
$$= x_\mu\partial_\nu - x_\mu x_\nu(x\cdot\partial - \sigma) - \frac{\sigma}{n}\delta_{\mu\nu}, \quad (2.5)$$

where $x\cdot\partial \equiv x_\mu\partial_\mu$, $\sigma = \frac{1}{2}(-n + \tau)$ and where we have taken the x_μ to be the Cartesian coordinates on the $(n-1)$-dimensional real sphere S_{n-1}, i.e., $x^2 = 1$. One then obtains the Lie algebra in the form

$$[X_{\mu\nu}, X_{\rho\sigma}] = \delta_{\nu\rho}X_{\mu\sigma} - \delta_{\mu\sigma}X_{\rho\nu}, \quad (2.6)$$

and $\mathrm{Tr}X = 0$. By identifying the $(n-1)$ independent commuting $X_{\mu\mu}$'s (no sum) as the Cartan subalgebra and the $X_{\mu\nu}$ $(\mu \neq \nu)$ as the root vectors, one identifies[14] the Cartan class A_{n-1}.

In order to see what type of representations are allowed in our constructions, we notice the following relation in the enveloping algebra of $sl(n,\mathbb{R})$,

$$X_{\mu\lambda}X_{\lambda\nu} = X_{\mu\nu}[(n+\sigma)(n-1) - \sigma]/n + \delta_{\mu\nu}(n-1)\sigma(\sigma+n)/n^2, \quad (2.7)$$

which can be used to express all the higher-order Casimir operators in terms of the second-order operator $C_2 = X_{\mu\nu}X_{\nu\mu}$; hence we have only a most degenerate series of representation. This reflects the fact that our representations are built on a rank one homogeneous space S_{n-1}. Contracting over μ and ν in (2.7), we have $C_2 = (n-1)\sigma(\sigma+n)/n$.

The fact that we have an $sl(n,\mathbb{R})$ form of A_{n-1} is indicated by Eq. (2.6) and the form (2.5) with the specification of the hermiticity properties which must await the introduction of a Hilbert space structure which will be discussed in Sec. 3. Suffice it now to say that all the generators $X_{\mu\nu}$ will be anti-Hermitian under the usual scalar product on the sphere S_{n-1} with the choice $\sigma = -\frac{1}{2}n + i\rho$, ρ real, i.e., for $\tau = 2i\rho$. Indeed, with this choice of σ, C_2 is $-n(n-1)/4 - (n-1)\rho^2/n$. These representations are reducible as can be seen from the fact that the generators $X_{\mu\nu}$ are all even functions of x_μ. An extra parity label ϵ is thus needed to specify irreducible representations. Then we can say that the parameters (ρ, ϵ) label the representations of a principal most degenerate series of $sl(n,\mathbb{R})$ built on the space of square-integrable functions on the sphere.

B. $i_2\, u(n) \Rightarrow sl(n,\mathbb{C}) \oplus u(1)$

In analogy with the previous section, we consider the algebra $u(n)$, the usual metric-preserving algebra for the complex space $\mathbb{C}^n$ and adjoin an ideal formed by the set of commuting second-rank mixed tensors $Z_{\mu\nu}$ with the symmetry property $\bar{Z}_{\mu\nu} = Z_{\nu\mu}$ (the bar denotes complex conjugation). The $i_2u(n)$ algebra is then defined through the commutation relations[14]

$$[C_{\mu\nu}, C_{\rho\sigma}] = \delta_{\nu\rho}C_{\mu\sigma} - \delta_{\mu\sigma}C_{\rho\nu}, \quad (2.8a)$$

$$[C_{\mu\nu}, Z_{\rho\sigma}] = \delta_{\nu\rho}Z_{\mu\sigma} - \delta_{\mu\sigma}Z_{\rho\nu}, \quad (2.8b)$$

and the two Z's commute.

The extension proposed in (2.3) is, for the unitary groups,

$$D_{\mu\nu} \equiv \tfrac{1}{4}[\Psi, Z_{\mu\nu}] + \tau Z_{\mu\nu}, \quad (2.9)$$

where $\Psi = -2C_{\mu\nu}C_{\nu\mu}$ is the $u(n)$ second-order Casimir invariant. Again we are unable to find an expansion for a completely general $Z_{\mu\nu}$ and again one does not have the freedom to add an additional vector space structure. The particular choice of representation for which the deformation can be carried through is[15] $C_{\mu\nu} = Z_\mu\bar{\partial}_\mu - \bar{z}_\nu\bar{\partial}_\mu$, $Z_{\mu\nu} = z_\mu\bar{z}_\nu/|z|^2$, $|z|^2 = z_\mu\bar{z}_\mu$, where we have used the notation $\partial_\mu \equiv \partial/\partial z_\mu$ and $\bar{\partial}_\mu \equiv \partial/\partial\bar{Z}_\mu$. It is then straightforward to verify that

$$[D_{\mu\nu}, D_{\rho\sigma}] = \delta_{\nu\rho}C_{\mu\sigma} - \delta_{\mu\sigma}C_{\rho\nu}, \quad (2.10)$$

while the D's obey the same transformation properties (2.8b) under the C's as the Z's. Moreover, the trace of $C_{\mu\nu}$ which we denote as $C \equiv C_{\mu\mu}(sum)$, provides a $u(1)$ subalgebra which not only commutes with the C's but also with the D's, thus providing the direct sum algebra $sl(n,\mathbb{C}) \oplus u(1)$. The existence of this $u(1)$ subalgebra arises from the fact that the generators $D_{\mu\nu}$ as well as each $|z_\mu|^2$, (μ fixed) remains invariant under $z_\mu \to e^{i\Phi}z_\mu$. This is the analog of the parity in the last section, and definite $u(1)$ transformation properties must be specified in order to get irreducible representations of $sl(n,\mathbb{C})$. Indeed, it will be seen shortly how this provides us with an additional Casimir operator.

A convenient form to display the $sl(n,\mathbb{C})$ structure is obtained by constructing the traceless combinations

$$X^\pm_{\mu\nu} \equiv \tfrac{1}{2}[C_{\mu\nu} \pm D_{\mu\nu} - \delta_{\mu\nu}(C \pm Tr D)/n], \quad (2.11a)$$

given explicitly by

$$X^+_{\mu\nu} = z_\mu\partial_\nu - \tfrac{1}{2}z_\mu\bar{z}_\nu(z\cdot\partial + \bar{z}\cdot\bar{\partial} - \sigma) - \frac{C+\sigma}{2n}\delta_{\mu\nu}, \quad (2.11b)$$

$$X^-_{\mu\nu} = -\bar{z}_\nu\bar{\partial}_\mu + \tfrac{1}{2}z_\mu\bar{z}_\nu(z\cdot\partial + \bar{z}\cdot\bar{\partial} - \sigma) - \frac{C-\sigma}{2n}\delta_{\mu\nu}, \quad (2.11c)$$

where $\sigma = -n + \tau$ and we have set $|z|^2 = 1$, so that the $(2n-1)$ independent real numbers in z are the complex Cartesian coordinates on the $(n-1)$-dimensional complex sphere $C_{n-1} \cong S_{2n-1}$. It is easy to check that all X^+'s commute with all X^-'s and hence we have explicitly a pair of commuting $sl(n,\mathbb{R})$ algebras given by (2.6). In this form the Cartan subalgebra is given by the $2(n-1)$ independent $X^\pm_{\mu\mu}$ (no sum) and one easily arrives at the Cartan structure $A_{n-1} \oplus A_{n-1}$. An additional advantage of the form (2.11) is the following convenient set of relations in the enveloping algebra of $sl(n,\mathbb{C})$

$$X^\pm_{\mu\lambda}X^\pm_{\lambda\nu} = \left[N^\pm \mp 1 + \frac{n-2}{2n}(C \pm \sigma)\right]X^\pm_{\mu\nu}$$
$$+ \frac{n-1}{2n}(C \pm \sigma)\left(\frac{C \pm \sigma}{n} \pm 1\right)\delta_{\mu\nu}, \quad (2.12)$$

where $N^+ = n$ and $N^- = 0$.

As in the last section, these can be used to express all the higher-order Casimir invariants in terms of the second-order ones[17]. We are thus led to a class of degenerate representations, but not just a most degenerate series: the two Casimir operators obtainable from (2.12) are $C_2^{\pm} \equiv X^+_{\mu\nu}X^+_{\nu\mu} \pm X^-_{\mu\nu}X^-_{\nu\mu}$ and take the values $(n-1)[\sigma(2n+\sigma)+C^2]/2n$ and $(n-1)C(n+\sigma)/n$, respectively.

Now using the fact that C is the generator of a $U(1)$ group and restricting the representations of this $U(1)$ to be single-valued, one finds that the eigenvalues of C are integers m. In the next section we shall introduce a definite scalar product on C_{n-1}, with respect to which hermiticity will be defined. For the generators (2.8a) we will have $C_{\mu\nu}{}^{\dagger} = C_{\nu\mu}$, while for the choice $\sigma = -n + i\rho$, (ρ real), i.e., τ imaginary in (2.9) $D_{\mu\nu}{}^{\dagger} = -D_{\mu\nu}$ and for (2.11), $(X^{\pm}_{\mu\nu})^{\dagger} = X^{\mp}_{\nu\mu}$. For this choice of σ and C, the Casimir operators $C_2^{\pm}$ are $-(m^2+4\rho^2)/2n - n/2$ and $2i\rho m/n$, respectively. Thus we have Hermitian representations of the $sl(n,\mathbb{C})$ algebra described by two numbers, a real ρ and an integer m.

In performing the previous deformation, we followed the analogy with the real sphere, making $x_\mu x_\nu \to z_\mu \bar{z}_\nu$ and using the metric-preserving algebra on the complex sphere. We could alternatively have decomposed $z_\mu \bar{z}_\nu$ into its real and imaginary parts and considered the corresponding deformations separately. Indeed, if we would have done this our σ would be $-n + \frac{1}{2}\tau$ instead of $-n+\tau$ making it more compatible with both the real and quarternionic cases. In the next section, when we consider the corresponding quaternionic case, it will be expedient to work in terms of real components due to the noncommutativity of the quaternions themselves. We shall indicate there the corresponding restrictions which yield the $sl(n,\mathbb{C})$, $sl(n,\mathbb{R})$, $u(n)$, and $so(n)$ subalgebras.

C. $i_2 sp(n) \oplus sp(1) \Rightarrow sl(n,\mathbb{Q}) \oplus sp(1)$

Since the symplectic algebra $sp(n)$ is the metric-preserving algebra for the n-dimensional quaternionic plane[18] $\mathbb{Q}^n$, it seems natural to carry the analogy with the last two sections one step further and look for the corresponding deformation to essentially $sl(n,\mathbb{Q})$, the special linear algebra over the noncommutative quaternionic field[14] (continuous division ring) $\mathbb{Q}$. Since the quaternions are perhaps not so well known, we present first a brief review of their properties[19].

The quaternions form a four-dimensional noncommutative algebra over the field of real numbers with a base composed of $\mathbf{e}_\alpha$ ($\alpha = 0,1,2,3$) whose multiplication table is

$$\mathbf{e}_0\mathbf{e}_i = \mathbf{e}_i\mathbf{e}_0 = \mathbf{e}_i, \qquad \mathbf{e}_i{}^2 = -\mathbf{e}_0, \mathbf{e}_i\mathbf{e}_j = \epsilon_{ijk}\mathbf{e}_k, \tag{2.13}$$

where $i,j,k = 1,2,3$. We use the convention that the early Greek letters range from 0 to 3, whereas the middle Latin letters over 1, 2, 3, reserving the middle Greek letters for the tensor indices. A quaternion can thus be written as $\mathbf{q} = q^\alpha \mathbf{e}_\alpha$. The quaternionic conjugate is defined as $\mathbf{q}^* \equiv q^0\mathbf{e}_0 - q^i\mathbf{e}_i$ and one verifies that $|\mathbf{q}|^2 \equiv \mathbf{q}^*\mathbf{q} = \mathbf{q}\mathbf{q}^* = q^\alpha q^\alpha$ is a real nonnegative number which vanishes iff $\mathbf{q}$ itself vanishes. We can form the quaternionic n plane $\mathbb{Q}^n$ by taking the n-fold tensor product of $\mathbb{Q}$, which forms a vector space endowed with a scalar product $\mathbb{Q}^n \times \mathbb{Q}^n \to \mathbb{Q}$ given by $\mathbf{u}\cdot\mathbf{q} \equiv \mathbf{u}^*_\mu\mathbf{q}_\mu$. The norm induced by this scalar product is $|\mathbf{q}|^2 \equiv \mathbf{q}\cdot\mathbf{q} = \mathbf{q}^*_\mu\mathbf{q}_\mu = \mathbf{q}_\mu\mathbf{q}^*_\mu = q^\alpha_\mu q^\alpha_\mu$. The scalar product defined above is left invariant by the group of n-dimensional symplectic transformations whose infinitesimal generators $M^\alpha_{\mu\nu}$ can be given in two different ways depending on whether the group action is defined from the right or from the left. This will be detailed in Sec. 3. An arbitrary second-rank mixed tensor with components $Q^\alpha_{\mu\nu}$ transforms under the generators of $sp(n)$ as

$$[M^0_{\mu\nu}, Q^\alpha_{\rho\sigma}] = \delta_{\nu\rho}Q^\alpha_{\mu\sigma} - \delta_{\mu\rho}Q^\alpha_{\nu\sigma} + \delta_{\nu\sigma}Q^\alpha_{\rho\mu} - \delta_{\mu\sigma}Q^\alpha_{\rho\nu}, \tag{2.14a}$$

$$[M^i_{\mu\nu}, Q^0_{\rho\sigma}] = \delta_{\nu\rho}Q^i_{\mu\sigma} - \delta_{\mu\rho}Q^i_{\nu\sigma} - \delta_{\nu\sigma}Q^i_{\rho\mu} + \delta_{\mu\sigma}Q^i_{\rho\nu}, \tag{2.14b}$$

$$[M^i_{\mu\nu}, Q^i_{\rho\sigma}] = -\delta_{\nu\rho}Q^0_{\mu\sigma} - \delta_{\mu\rho}Q^0_{\nu\sigma} + \delta_{\nu\sigma}Q^0_{\rho\mu} + \delta_{\mu\sigma}Q^0_{\rho\nu} \quad \text{(no sum on } i\text{)} \tag{2.14c}$$

$$[M^i_{\mu\nu}, Q^j_{\rho\sigma}] = \epsilon_{ijk}(\delta_{\nu\rho}Q^k_{\mu\sigma} + \delta_{\mu\rho}Q^k_{\nu\sigma} + \delta_{\nu\sigma}Q^k_{\rho\mu} + \delta_{\mu\sigma}Q^k_{\rho\nu}). \tag{2.14d}$$

The symplectic algebra[14] $sp(n)$ spanned by the M's satisfies (2.14) with the additional specification that $M^0_{\mu\nu} = -M^0_{\nu\mu}$ and $M^i_{\mu\nu} = M^i_{\nu\mu}$. We can realize this algebra[15] on $\mathbb{Q}^n$ as

$$M^0_{\mu\nu} = q^\alpha_\mu\partial^\alpha_\nu - q^\alpha_\nu\partial^\alpha_\mu, \tag{2.15a}$$

$$M^{i\pm}_{\mu\nu} = \pm(q^i_\mu\partial^0_\nu + q^i_\nu\partial^0_\mu - q^0_\mu\partial^i_\nu - q^0_\nu\partial^i_\mu) - \epsilon_{ijk}(q^j_\mu\partial^k_\nu + q^j_\nu\partial^k_\mu), \tag{2.15b}$$

where $\partial^\alpha_\mu \equiv \partial/\partial q^\alpha_\mu$ and where (+) or (−) designates action from the left or right with respect to quaternionic multiplication. One sees then that $\mathbf{q}_\rho\mathbf{q}^*_\sigma$ transforms as a mixed second-rank tensor under the M^+'s but not under the M^-'s. Conversely, the quaternionic conjugate $\mathbf{q}^*_\sigma\mathbf{q}_\rho$ transforms as a mixed second-rank tensor under the M^-'s only. Furthermore, the commutator of the M^+'s with the M^-'s does not close to form a Lie algebra of finite dimension. It can be seen, however, that the traces

$$M^{i\mp} \equiv \tfrac{1}{4}M^{i\mp}_{\mu\mu} = \pm\tfrac{1}{2}(q^0_\mu\partial^i_\mu - q^i_\mu\partial^0_\mu) - \tfrac{1}{2}\epsilon_{ijk}q^j_\mu\partial^k_\mu, \tag{2.15c}$$

commute with the $M^{i\pm}_{\mu\nu}$'s, each forming the algebra $sp(1)^{\mp} \cong su(2)$. Moreover, M^{i-} commutes with $\mathbf{q}_\rho\mathbf{q}^*_\sigma$ and M^{i+} commutes with $\mathbf{q}^*_\sigma\mathbf{q}_\rho$. Hence, we finally arrive at two possible isomorphic algebras which we denote as $i_2sp(n)^+ \oplus sp(1)^-$ and $i_2sp(n)^- \oplus sp(1)^+$ spanned by $\{M^{\alpha+}_{\mu\nu}, M^{i-}, \mathbf{q}_\mu\mathbf{q}^*_\nu\}$ and $\{M^{\alpha-}_{\mu\nu}, M^{i+}, \mathbf{q}^*_\mu\mathbf{q}_\nu\}$, respectively.

Using either of these algebras we are now in a position to write down the deformation formulas which are a generalization of Eqs. (2.3) and (2.9) to the quaternionic case. We use the first set of operators spanning $i_2sp(n)^+ \oplus sp(1)^-$ and consider

$$N^\alpha_{\mu\nu} \equiv \tfrac{1}{2}[\Omega^+, (\mathbf{q}_\nu\mathbf{q}^*_\mu)^\alpha/|\mathbf{q}|^2] + \tau(\mathbf{q}_\nu\mathbf{q}^*_\mu)^\alpha/|\mathbf{q}|^2, \tag{2.16}$$

where $\Omega^\pm = -\frac{1}{2}M^0_{\mu\nu}M^0_{\nu\mu} + \frac{1}{2}M^{i\pm}_{\mu\nu}M^{i\pm}_{\nu\mu}$ is the second-order Casimir invariant for $sp(n)^\pm$. If we consider the combinations

$$X^0_{\mu\nu} \equiv \frac{1}{2}(M^0_{\mu\nu} + N^0_{\mu\nu}) - \frac{1}{2n}\delta_{\mu\nu}\,\mathrm{Tr}N^0, \tag{2.17a}$$

$$X^i_{\mu\nu} \equiv \tfrac{1}{2}(M^i_{\mu\nu} + N^i_{\mu\nu}), \tag{2.17b}$$

where the X^0 are built to be traceless, and place $|\mathbf{q}|^2 = 1$, the $(n-1)$-dimensional quaternionic sphere $Q_{n-1} \cong S_{4n-1}$ where the $4n-1$ independent real q^α_μ are the Cartesian coordinates, and set $\sigma = -2n + \frac{1}{2}\tau$, we arrive, after a fairly tedious calculation, at the explicit form

$$X^0_{\mu\nu} = q^\alpha_\mu\partial^\alpha_\nu - (\mathbf{q}_\nu\mathbf{q}^*_\mu)^0(\mathbf{q}\cdot\partial - \sigma) - \delta_{\mu\nu}\sigma/n, \tag{2.18a}$$

$$X^i_{\mu\nu} = q^i_\mu \partial^0_\nu - q^0_\mu \partial^i_\nu - \epsilon_{ijk} q^j_\mu \partial^k_\nu - (\mathbf{q}_\nu \mathbf{q}^*_\mu)^i (\mathbf{q}\cdot\partial - \sigma), \quad \mathbf{q}\cdot\partial \equiv q^\alpha_\mu \partial^\alpha_\mu. \tag{2.18b}$$

It can be checked that the X's generate the Lie algebra[14]

$$[X^0_{\mu\nu}, X^\alpha_{\rho\sigma}] = \delta_{\nu\rho} X^\alpha_{\mu\sigma} - \delta_{\mu\sigma} X^\alpha_{\rho\nu}, \tag{2.19a}$$

$$[X^i_{\mu\nu}, X^i_{\rho\sigma}] = -\delta_{\nu\rho} X^0_{\mu\sigma} + \delta_{\mu\sigma} X^0_{\rho\nu}, \tag{2.19b}$$

$$[X^i_{\mu\nu}, X^j_{\rho\sigma}] = \epsilon_{ijk}(\delta_{\nu\rho} X^k_{\mu\sigma} + \delta_{\mu\sigma} X^k_{\rho\nu}), \tag{2.19c}$$

identified as $sl(n, \mathbb{Q})$.

It is not difficult to see that the X^0's span an $sl(n, \mathbb{R})$ subalgebra while the X^0's and X^i's for one fixed i, span an $sl(n, \mathbb{C}) \oplus u(1)$ subalgebra. By taking the $X^\alpha_{\mu\nu}$'s given by Eq. (2.18) and the traces M^{i-} in (2.15c), we obtain an algebra $sl(n, \mathbb{Q})^+ \oplus sp(1)^-$. It is easy to check that, indeed, M^{i-} commutes with all the $X^{\alpha+}_{\mu\nu}$'s. Alternatively, we could have constructed the algebra $sl(n, \mathbb{Q}) - \oplus sp(1)^+$ by starting from the $i_2 sp(n)^- \oplus sp(1)^+$ algebra. The net effect of this on Eq. (2.18) is to reverse the sign of the non-epsilon terms in the expression for the $X^i_{\mu\nu}$'s. The $sl(n, \mathbb{Q})$ structure of Eqs. (2.19) can be brought out by taking $X^0_{\mu\mu}$ and, say, $X^1_{\mu\mu}$ (no sum) as the Cartan subalgebra. The root vectors are then given by $X^0_{\mu\nu} \pm iX^1_{\mu\nu}$ and $X^2_{\mu\nu} \pm iX^3_{\mu\nu}$. This identifies the Cartan class A_{2n-1}.

The role played by the $sp(1)$ is analogous to that of $u(1)$ in the complex case. Both the set of generators $X^\alpha_{\mu\nu}$ and each term $|\mathbf{q}_\mu|^2$ of the quaternionic sphere are invariant under $sp(1)^-$. Again, definite $sp(1)^-$ transformation properties must be specified in order to get irreducible representations of $sl(n, \mathbb{Q})^+$. This will become clearer in the group theoretical context in the following section.

In the enveloping algebra of $sl(n, \mathbb{Q})$ we were able to derive one analog of Eqs. (2.7) and (2.12)

$$X^0_{\mu\lambda} X^0_{\lambda\nu} - X^{i+}_{\mu\lambda} X^{i+}_{\lambda\nu} = -X^{i-}_{\mu\nu} M^{i-} + \left[4(n-1) + \frac{n-2}{n}\right] X^0_{\mu\nu} + \frac{n-1}{n^2}\sigma(\sigma + 4n)\delta_{\mu\nu}. \tag{2.20}$$

It seems, however, that this relation is not by itself sufficient to reduce all higher-order Casimir operators to the second-order one. Indeed, we expect more non-independent Casimir invariants due to the existence of the $sp(1)^-$ algebra. These invariants will be of higher order than second, since in contrast to the $sl(n, \mathbb{C})$ case, $sl(n, \mathbb{Q})$ is a simple real Lie algebra. Due to the complexity in deriving such relations, however, we have thus far been unable to find them. Relation (2.20) does provide the second-order Casimir invariant $C_2 \equiv X^0_{\mu\nu} X^0_{\nu\mu} - X^{i+}_{\mu\nu} X^{i+}_{\nu\mu} = -M^{i-} M^{i-} + (n-1)\sigma(\sigma + 4n)/n$.

The $sp(1)^-$ invariant $(M^-)^2 \equiv M^{i-} M^{i-}$ can be chosen to define a basis where its eigenvalues are $l(l+1)$ (l integer on half-integer). In the next section we shall introduce a definite scalar product on Q_{n-1}, with respect to which all the operators used in this section are anti-Hermitian if we choose $\sigma = -2n + i\rho$ (ρ real). For these values of σ, the eigenvalue of the Casimir invariant C_2 is the real number $-l(l+1) - (n-1)(4n^2 + \rho^2)/n$.

3. HOMOGENEOUS FUNCTIONS AND MULTIPLIER REPRESENTATIONS

In this section we shall relate the expressions obtained in the previous sections by the deformation of inhomogeneous algebras to the corresponding technique of constructing multiplier representations for the group from certain classes of homogeneous functions similar to those introduced by Bargmann[10] and Gel'fand.[11] Indeed, from the expressions for the generators given by Eqs. (2.5), (2.11), and (2.18), the terms in σ indicate that they should upon integration give rise to multiplier representations.[8] Rather than exponentiate these expressions directly, however, we prefer to construct the multipliers by Gel'fand's method of homogeneous functions[11] and obtain the corresponding generators for the one-parameter subgroups. This procedure lends insight into the expansions of the form (2.3) on the global group level. All the known expansions of the form (2.3) display this correspondence to homogeneous functions.[8]

It is not difficult to see that the spheres S_{n-1}, C_{n-1}, and Q_{n-1} introduced in the last section correspond to homogeneous spaces of the groups $SL(n, \mathbb{F})$ of $n \times n$ matrices $G = \|g_{\mu\nu}\|$, $(\mu, \nu = 1, \ldots, n)$, $\det G = 1, g_{\mu\nu} \in \mathbb{F}$, where $\mathbb{F}$ indicates the real, complex and quaternion fields. Indeed, consider the Iwasawa decomposition[4] of $SL(n, \mathbb{F}) = KAN$ where K is $SO(n)$, $SU(n)$, and $Sp(n)$, respectively, A is the $(n-1)$-dimensional Abelian subgroup of diagonal matrices of $SL(n, \mathbb{F})$, and N is the nilpotent subgroup of lower-triangular matrices. Then in each case, if K' is the canonical subgroup $SO(n-1)$, $SU(n-1)$, and $Sp(n-1)$, respectively. $K'AN$ is the subgroup of $n \times n$ matrices $G' = \|g'_{\mu\nu}\|$ such that the elements $g'_{in} = 0$ $(i = 1, \ldots, n-1)$, and $\det G' = 1$. The homogeneous spaces $\Omega \equiv KAN/K'AN$ are then, respectively, the spheres S_{n-1}, C_{n-1}, and Q_{n-1}. The action of the group element $g \in G$ on the Cartesian coordinates $s_\mu (\mu = 1, \ldots, n)$, $s_\mu \in \mathbb{F}$, $s^*_\mu s_{\mu=1}$ from the *left* is given by

$$s_\mu \xrightarrow{gL} s'_\mu = \frac{r}{r'} g^{-1}_{\mu\nu} s_\nu, \quad \frac{r}{r'} = [s^*_\lambda g^{-1*}_{\rho\lambda} g^{-1}_{\rho\sigma} s_\sigma]^{-1/2}, \tag{3.1a}$$

and if the field $\mathbb{F}$ is $\mathbb{Q}$ we can also have a distinct action from the *right* as given by

$$s_\mu \xrightarrow{gR} s''_\mu = \frac{r}{r''} s_\nu g^{-1*}_{\mu\nu}, \quad \frac{r}{r''} = [g^{-1}_{\rho\lambda} s^*_\lambda s_\sigma g^{-1*}_{\rho\sigma}]^{-1/2}, \tag{3.1b}$$

where it should be understood that the involutive automorphism *: $s_\mu \to s_\mu{}^*$ is the identity for $\mathbb{F} = \mathbb{R}$, complex conjugation for $\mathbb{F} = \mathbb{C}$, and quaternionic conjugation for $\mathbb{F} = \mathbb{Q}$. The subgroup $G' = K'AN$ is then the stability group of the point $(s_\mu) = (0, \ldots, 0, 1)$ on Ω. The transformations of K [$SO(n)$, $SU(n)$, and $Sp(n)$, respectively] are the largest group of *rigid* transformations of the sphere Ω since they leave the measure $d\Omega$ on the sphere invariant. The rest of the transformations $g \in G$ will produce a "*deformation*" of the surface of Ω, where the Jacobian is

$$(d\Omega/d\Omega')^L = (r'/r)^p, \tag{3.2a}$$

$$(d\Omega/d\Omega'')^R = (r''/r)^p, \tag{3.2b}$$

where $p = n$ dim$\mathbb{F}$, i.e., $p = n, 2n$, and $4n$ for $\mathbb{F} = \mathbb{R}, \mathbb{C}$, and $\mathbb{Q}$, respectively. In the former two cases, (3.2a) and (3.2b) are equal.

In comparing this approach with the one used for[4,8,20] $SO(n, 1) \supset SO(n)$, $SU(n, 1) \supset SU(n)$, and[21] $Sp(n, 1) \supset Sp(n)$, we notice that there is one essential difference with the above and that is that in the case of these groups the subgroup K' is the centralizer of A and the normalizer of N in K. This has the consequence that the irreducible representations of the subgroup $K'AN$ are just direct products of irreducible representations

of K' and the irreducible representations (characters) of A. Hence, one can induce all these representations to the full group KAN. In the case of $SL(n,\mathbb{F})$, however, K' is no longer the centralizer of A and only a "most degenerate" representation of $K'AN$ labeled by a character of A can be induced[22] to irreducible representations of $SL(n,\mathbb{F})$. These are just the representations described in the previous sections by the deformation of the corresponding representations of the inhomogeneous Lie algebras.

A. $SL(n,\mathbb{R})$

Consider the space of homogeneous functions over the n-dimensional real plane $\mathbb{R}^n$ which are infinitely differentiable (except possibly at the origin) and satisfy[11,23]

$$F(ay_\mu) = a^\sigma \operatorname{sgn}^\epsilon a F(y_\mu), \tag{3.3}$$

where $a, y_\mu \in \mathbb{R}$ and $\epsilon = 0, 1$. Now representations of $SL(n,\mathbb{R})$ can be constructed over this space as representations by left action

$$\tilde{T}_g F(y_\mu) = F(y'_\mu) = F(g^{-1}_{\mu\nu} y_\nu), \qquad g \in SL(n,\mathbb{R}). \tag{3.4}$$

Since the functions F satisfy (3.3), i.e., are homogeneous functions, the representation (3.4) gives rise to a representation on the unit sphere S_{n-1} in the following way: From (3.3) we see that we can define a function on S_{n-1} through $F(y_\mu) = r^\sigma f(x_\mu)$, with $x_\mu \in S_{n-1}$ and $r \geq 0$. A simple calculation shows that (3.4) induces the representation

$$T^\sigma_g f(x_\mu) = (r'/r)^\sigma f(x'_\mu) \tag{3.5}$$

over functions on S_{n-1}, where $T^\sigma_g = r^{-\sigma}\tilde{T}_g r^\sigma$ and the group action is given by (3.1a). Furthermore, from (3.4) and the infinite differentiability of the F's, it follows that the f's span the space $\mathfrak{D}^\epsilon$ of infinitely differentiable functions on S_{n-1} which satisfy

$$f(-x_\mu) = (-i)^\epsilon f(x_\mu). \tag{3.6}$$

The function $(r'/r)^\sigma$ is a *multiplier* which trivially satisfies the condition[10,11] $(r''/r')^\sigma (r'/r)^\sigma = (r''/r)^\sigma$ and hence Eq. (3.5) is indeed a representation of $SL(n,\mathbb{R})$.

We obtain the infinitesimal generators of $SL(n,\mathbb{R})$ by considering the one-parameter subgroups $g_{\mu\nu}(t)$ which to first order are $g_{\nu\mu}(t) \simeq \delta_{\nu\mu} - t\alpha_{\nu\mu}$, $T^\sigma_g \simeq \mathbf{1} + t\alpha_{\nu\mu} X_{\mu\nu}$, where we leave implicit the dependence of $X_{\mu\nu}$ on σ. As it is well known that demanding $\det\| g_{\mu\nu}\| = 1$ imposes the tracelessness conditions on the generators. We can use (3.1a) and (3.5) to arrive exactly at the generators (2.5) of $sl(n,\mathbb{R})$ obtained in Sec. 2A.

One can obtain unitary representations of $SL(n,\mathbb{R})$ over S_{n-1} taking the vector spaces $\mathfrak{D}^\epsilon$, $\epsilon = 0, 1$, and completing them with respect to the norm induced by the inner product

$$(f^\epsilon_1, f^\epsilon_2)_S = \int d\Omega(x) \overline{f^\epsilon_1(x)} f^\epsilon_2(x), \tag{3.7}$$

where $d\Omega(x)$ is the $SO(n)$-invariant measure on S_{n-1}. Then one can see that the representation (3.6) is unitary with respect to (3.7) with $\sigma = -\frac{1}{2}n + i\rho$ (ρ real). The multiplier in (3.5) is just what is needed to offset the transformation (3.2a) of the measure $d\Omega(x)$ under $SL(n,\mathbb{R})$.

B. $SL(n,\mathbb{C})$

Consider the space of functions $F(\omega_\mu)$ over the complex n plane $\mathbb{C}^n$, infinitely differentiable in ω_μ and $\overline{\omega}_\mu$ (except possibly at the origin), which satisfy[11,24]

$$F(a\omega_\mu) = a^{\sigma_1}\bar{a}^{\sigma_2} F(\omega_\mu), \tag{3.8}$$

where $a, \omega_\mu, \sigma_1, \sigma_2 \in \mathbb{C}$. Furthermore, we note that $F(e^{i\psi}\omega_\mu) = \exp[i(\sigma_1 - \sigma_2)\psi]F(\omega_\mu)$, thus providing a representation of $U(1)$. Requiring this representation to be single-valued implies that $\sigma_1 - \sigma_2 = m$ is an integer. Then the functions F are said to be *homogeneous* of degree (σ, m) where $\sigma = \sigma_1 + \sigma_2$. Now representations of $SL(n,\mathbb{C})$ can be constructed through left action as

$$\tilde{T}_g F(\omega_\mu) = F(\omega'_\mu) = F(g^{-1}_{\mu\nu}\omega_\nu), \qquad g \in SL(n,\mathbb{C}). \tag{3.9}$$

The homogeneity of the functions $F(\omega)$ allows us to construct functions over C_{n-1} as $F(\omega_\mu) = r^\sigma f(z_\mu)$ with $z \in C_{n-1}$, $r \geq 0$. The representation (3.9) induces the multiplier representation

$$T^\sigma_g f(z_\mu) = (r'/r)^\sigma f(z'_\mu) \tag{3.10}$$

over functions on C_{n-1}, where $T^\sigma_g = r^{-\sigma}\tilde{T}_g r^\sigma$ and (3.1a) for C_{n-1}. It then follows that the functions $f(z)$ are infinitely differentiable in z_μ and $\bar{z}_\mu$ with the auxiliary condition

$$f(e^{i\psi}z_\mu) = e^{im\psi} f(z_\mu). \tag{3.11}$$

We denote this space of functions as $\mathfrak{D}^m$. Actually (3.11) defines a representation of the $U(1)$ subgroup of $SL(n,\mathbb{C}) \otimes U(1)$ as $T_\mu f(z_\mu) = f(e^{i\psi}z_\mu) = e^{im\psi} f(z_\mu)$.

In the same way as in the preceding section, the infinitesimal generators of both the representation of $SL(n,\mathbb{C})$ (3.10) and $U(1)$ (3.11) can be found with the parametrization

$$u \simeq 1 - i\psi, \qquad T_u \simeq \mathbf{1} + i\psi C,$$

$$g_{\nu\mu}(t) \simeq \delta_{\nu\mu} - t\alpha_{\nu\mu}, \qquad T^\sigma_g \simeq \mathbf{1} + t(\alpha_{\nu\mu}X^+_{\mu\nu} - \overline{\alpha_{\nu\mu}}X^-_{\mu\nu}),$$

By imposing the condition of tracelessness on these generators, we arrive at the expressions for $X^\pm_{\mu\nu}$ given by Eq. (2.11).

We endow the spaces $\mathfrak{D}^m$ with a Hilbert space structure by completion with respect to the norm induced by the inner product

$$(f^m_1, f^m_2)_C = \int d\Omega(z) \overline{f^m_1(z)} f^m_2(z), \tag{3.12}$$

where $d\Omega(z)$ is the $U(n)$-invariant measure on C_{n-1}. It follows that the representation (3.10) will be unitary with respect to the inner product (3.12) if we choose $\sigma = -n + i\rho$ (ρ real), since the multiplier just cancels the change in the measure (3.2a).

C. $SL(n,\mathbb{Q})$

The description of $SL(n,\mathbb{Q})$ follows those of $SL(n,\mathbb{R})$ and $SL(n,\mathbb{C})$, the major difference now being that the multiplication of quaternions in the representation can be taken from the left or right, giving rise to two different realizations of $SL(n,\mathbb{Q})$. Let $F(\mathbf{u}_\mu)$ be an infinitely differentiable complex-valued function on the quaternionic n plane $\mathbb{Q}^n$ (except possibly at the origin). We can define representations by left and right group action as

$$\tilde{T}^L_g F(\mathbf{u}_\mu) = F(\mathbf{g}^{-1}_{\mu\nu}\mathbf{u}_\nu), \tag{3.13a}$$

$$\tilde{T}^R_g F(\mathbf{u}_\mu) = F(\mathbf{u}_\nu \mathbf{g}^{-1*}_{\mu\nu}), \tag{3.13b}$$

with $\mathbf{u}_\mu \in \mathbb{Q}$, $g \in SL(n,\mathbb{Q})$. Notice that we always have left multiplication with respect to the tensor indices. As

in the previous cases, we want to restrict our class of functions F to be homogeneous functions of $\mathbb{Q}^n$ in some sense. Due to the quaternion noncommutativity, there is an ambiguity in factoring out quaternions as done in Eq. (3.8) for $\mathbb{C}$. We thus consider "*homogeneous*" functions (in an expanded sense) of degree (σ, l, m) which satisfy

$$F^l_m(a\mathbf{u}_\mu) = a^\sigma F^l_m(\mathbf{u}_\mu), \qquad a \in \mathbb{R}, \tag{3.14a}$$

$$F^l_m(\mathbf{u}_\mu \mathbf{s}) = \sum_{m'} F^l_{m'}(\mathbf{u}_\mu) D^l_{m'm}(\mathbf{s}(\alpha,\beta,\gamma)), \tag{3.14b}$$

where we have used the familiar Wigner D function for $SU(2) \cong Sp(1)$, and $\mathbf{s}$ is a unit quaternion $|\mathbf{s}|^2 = 1$, parametrized by Euler angles $\mathbf{s}(\alpha,\beta,\gamma) = \exp(\mathbf{e}_3\alpha)\exp(\mathbf{e}_2\beta)\exp(\mathbf{e}_3\gamma)$.

We recognize that in order to write an equations such as (3.14b) we must consider vector-valued functions $F^l(\mathbf{u}_\mu)$ on $\mathbb{Q}^n$ of degree l. There is an expression analogous to (3.14b) obtained by multiplication from the left. As in previous sections, we construct functions f on Q_{n-1} through $F^l_m(\mathbf{u}_\mu) = r^\sigma f^l_m(\mathbf{q}_\mu)$, $\mathbf{q} \in Q_{n-1}$. These will constitute the space $\mathfrak{D}^{l,m}$ of infinitely differentiable functions over Q_{n-1}. We then construct a multiplier representation of $SL(n, \mathbb{Q})^L$ on $\mathfrak{D}^{l,m}$ as

$$T^{L\sigma}_g f^l_m(\mathbf{q}_\mu) = (r'/r)^\sigma f^l_m(\mathbf{q}'_\mu), \tag{3.15}$$

where $T^{L\sigma}_g = r^{-\sigma} T^L_g r^\sigma$ and $\mathbf{q}'_\mu$ is found in (3.1a). There is an expression similar to (3.17) for T^R_g by using (3.1b). Now Eq. (3.14b) simply becomes

$$f^l_m(\mathbf{q}_\mu \mathbf{s}) = \sum_{m'} f^l_{m'}(\mathbf{q}_\mu) D^l_{m'm}(\mathbf{s}). \tag{3.16}$$

Notice that $\mathfrak{D}^{lm}$ is not invariant under $Sp(1)$. Indeed, it is seen that the f^l_m transform as the components of a rank l spherical tensor under $Sp(1)^R$ acting from the right. This equation also defines a representation of $Sp(1)^R$ by right action, i.e., $T^R_{\mathbf{s}} f^l_m(\mathbf{q}_\mu) = f^l_m(\mathbf{q}_\mu \mathbf{s})$. It can be shown easily that this action and (3.15) commute, leading to the structure $SL(n, \mathbb{Q})^L \otimes Sp(1)^R$. It is clear that one can similarly construct $SL(n, \mathbb{Q})^R \otimes Sp(1)^L$.

One can then obtain the infinitesimal generators by using the parametrization $\mathbf{g}_{\mu\nu} = \delta_{\mu\nu}\mathbf{e}_0 - t\boldsymbol{\alpha}_{\mu\nu}$ and imposing the condition of tracelessness on the $\alpha^0_{\mu\nu}$ term to arrive at the generators given in (2.15). We can similarly obtain the infinitesimal generators of T^R_g by reversing the sign of all non-epsilon terms in the coefficient of $\alpha^i_{\mu\nu}$. Also from Eq. (3.16) we can obtain the generators of $Sp(1)^R$; they are the traces M^{i-} in (2.15c).

We introduce the Hilbert space structure by completing $\mathfrak{D}^{lm}$ with respect to the norm induced by the inner product

$$(f^l_{1m}, f^l_{2m})_Q = \int d\Omega(\mathbf{q}) \overline{f^l_{1m}(\mathbf{q})} f^l_{2m}(\mathbf{q}), \tag{3.17}$$

where $\mathbf{q} \in Q_{n-1}$ and $d\Omega(\mathbf{q})$ is the $Sp(n)^L \otimes Sp(1)^R$ invariant measure on Q_{n-1}. Notice that there is no sum over m here since the space $\mathfrak{D}^{l,m}$ is invariant under the representation (3.15) of $SL(n, \mathbb{Q})^L$. This representation is unitary if we choose $\sigma = -2n + i\rho$ due to the $SL(n, \mathbb{Q})$ transformation of $d\Omega(\mathbf{q})$ in (3.2a).

Also it can be seen that the representation (3.16) of $Sp(1)^R$ is unitary for l integer or half-integer, upon introduction of the usual vector space inner product

$$(f^l_1, f^l_2)(\mathbf{q}) = \sum_m \overline{f^l_{1m}(\mathbf{q})} f^l_{2m}(\mathbf{q}). \tag{3.18}$$

4. CONTRACTIONS OF REPRESENTATIONS

A. Of the algebra

The contraction[12] of the previous representations of the Lie algebras $sl(n, \mathbb{F}) \oplus a(\mathbb{F})$ is to the algebras $i_2 k(\mathbb{F}) \oplus a(\mathbb{F})$. We will adopt the notation used in the beginning of Sec. 3 and treat the three cases $\mathbb{F} = \mathbb{R}, \mathbb{C}$, and $\mathbb{Q}$ together, and hence let $k(\mathbb{F})$ denote, respectively, $so(n), u(n)$, and $sp(n)$, while $a(\mathbb{F})$ denotes $0, u(1)$, and $sp(1)$. The generators of $k(\mathbb{F})$ are $M^\alpha_{\mu\nu}$, where $\alpha = 0$ for $\mathbb{F} = \mathbb{R}$; $\alpha = 0, 1$ for $\mathbb{F} = \mathbb{C}$; and $\alpha = 0, 1, 2, 3$ for $\mathbb{F} = \mathbb{Q}$; the remaining generators are

$$N^\alpha_{\mu\nu} = \tfrac{1}{2}[\Omega, (s_\nu, s^*_\mu)^\alpha] + \tau(s_\nu s^*_\mu)^\alpha \tag{4.1}$$

with the same ranges of α. In order to perform the contraction one considers the generators $N^\alpha_{\mu\nu}/\tau$ as spanning along with the $M^\alpha_{\mu\nu}$ a sequence of representations denoted by $sl(n, \mathbb{F})_\tau$. Upon taking the limit as $|\tau| \to \infty$, one finds

$$\lim_{|\tau|\to\infty} \frac{1}{\tau} N^\alpha_{\mu\nu} = (s_\nu s^*_\mu)^\alpha, \tag{4.2}$$

whence we write $sl(n, \mathbb{F})_\tau \xrightarrow{|\tau|\to\infty} i_2 k(\mathbb{F})$. Equation (2.7), (2.12), and (2.20) become identities in the contraction limit. We note that no role is played by $a(\mathbb{F})$ in the contraction procedure. The deformation performed in Sec. 2 and the above contraction are inverse operation.[6] We make note that although our representations were built as deformations of $i_2 k(\mathbb{F}) \oplus a(\mathbb{F})$, they can also be viewed as expansions of the inhomogeneous algebra $ik(\mathbb{F}) \oplus a(\mathbb{F})$.

B. Of the group

The contraction of the corresponding group representations (3.5), (3.10), and (3.15) proceeds in the standard way[8,12] by allowing the group transformation $g(t)$ to approach the identity ($t = 0$) as we let $\rho \to \infty$ in the sequence of representations $SL(n, \mathbb{F})_{i\rho}$ in such a way that $t\rho = \xi$, a real constant. Thus we see from (3.1) that $s'_\mu \xrightarrow{t\to\infty} s_\mu$ and

$$\lim\left(\frac{r'}{r}\right)^{-p/2+i\rho} = \lim\left(1 + \frac{\xi}{\rho} s^*_\nu \frac{\alpha^*_{\mu\nu} + \alpha_{\nu\mu}}{2} s_\mu\right)^{i\rho}$$
$$= \exp[i\xi\alpha^\alpha_{\nu\mu}(s_\nu s^*_\mu)^\alpha], \tag{4.3}$$

where the same remarks for the cases $\mathbb{F} = \mathbb{R}, \mathbb{C}$, and $\mathbb{Q}$ apply. It should be noticed that only the symmetric part of $\alpha^0_{\nu\mu}$ and the antisymmetric one of $\alpha^i_{\nu\mu}$ contribute to the multiplier; the representations therefore contract as

$$\lim T^\sigma_g f(s_\mu) = \exp[i\xi\alpha^\alpha_{\nu\mu}(s_\nu s^*_\mu)^\alpha] f(s_\mu), \tag{4.4}$$

showing that only the "boost" group elements generated by (4.1) have a finite contraction limit. We thus have found the representations built in Sec. 3 to contract as $SL(n, \mathbb{R}) \to I_2 SO(n)$, $SL(n, \mathbb{C}) \otimes U(1) \to I_2 SU(n) \otimes U(1)$, and $SL(n, \mathbb{Q}) \otimes Sp(1) \to I_2 Sp(n) \otimes Sp(1)$.

5. CONCLUSION

We have exhibited deformations of inhomogeneizations of all the classical Cartan Lie algebras to those of the linear groups. These deformations are representation-dependent in that the procedure can be implemented only for representations which can be realized on rank one homogeneous spaces. While we have not been able to provide an 'if and only if' statement of this fact, we believe that we have indeed constructed

the most general representations from our deformation procedure. Thus while our family of deformations falls outside the class studied by Gilmore[9] (i.e., rank one coset space in the Cartan decomposition), it does so only in a mild way, since a rank one homogeneous space is involved. These representations are precisely those which do not exhibit multiplicity problems when reduced to the maximum compact subgroup. This supports a conjecture by Mukunda[25] and Hermann[4] that the ability to use the deformation or expansion algorithm is intimately connected with the nonexistence of multiplicity problems.

On the group level, it was shown that these deformations are related to multiplier representations and "deformations" of the homogeneous space. In this regard, there seems to be a need to establish a more thoroughgoing connection between the infinitesimal and global approaches. A related approach is to perform the deformations not merely on homogeneous spaces, but on the whole group manifold.[21] In this way, cases where multiplicity problems appear might be incorporated.

The multiplier representations discussed in this paper can be used to calculate the finite group element representation matrix elements[8] in the basis obtained by the canonical decomposition $SL(n, \mathbb{F}) \supset K(\mathbb{F})$ [$K = SO(n), SU(n)$, or $Sp(n)$]. Although we only explicitly constructed a principal series, other series (e.g., supplementary and discrete) should be obtainable by allowing a nonlocal measure[10,11] as has been done[26] for $SO(n, 1) \supset SO(n)$. It is also to be remarked that other noncompact chains can be discussed as well through our deformation procedure and multiplier representations implemented on hyperboloids $s_1 s_1^* + \cdots + s_k s_k^* - s_{k+1} s_{k+1}^* - \cdots - s_n s_n^* = 1$, as well as spheres. This would allow one to discuss such decompositions as $SL(n, \mathbb{R}) \supset SO(n-k, k)$, $SL(n, \mathbb{C}) \supset SU(n-k, k)$, and $SL(n, \mathbb{Q}) \supset Sp(n-k, k)$, $(k = 1, \ldots, n-1)$, without multiplicity problems beyond the doubling encountered in the reduction[8] $SO(n, 1) \supset SO(n-1, 1)$.

In conclusion it can be said that our realizations for $n = 2, SL(2, \mathbb{R}) \overset{2-1}{\cong} SO(2, 1)$, $SL(2, \mathbb{C}) \overset{2-1}{\cong} SO(3, 1)$ yield all the principal series representations and reproduce the known results on these groups by Bargmann[10] and Gel'fand and collaborators.[11] One can use similar procedures to discuss the representations of $SL(2, \mathbb{Q}) \overset{2-1}{\cong} SO(5, 1)$.

[1]S. Goshen and H. J. Lipkin, Ann. Phys. (N.Y.) **6**, 301 (1959); Ann. Phys. (N.Y.) **6**, 310 (1959).

[2]M. A. Melvin, Bull. Am. Phys. Soc. **7**, 493 (1962); Bull. Am. Phys. Soc. **8**, 356 (1963); A. Sankaranarayanan, Nuovo Cimento **38**, 1441 (1965); A. Sankaranarayanan and R. H. Good Jr., Phys. Rev. **B140**, 509 (1965); A. Bohm, Phys. Rev. **145**, 1212 (1966); E. Weimar, Nuovo Cimento Lett. **4**, 43 (1972).

[3]Y. Dothan, M. Gell-Mann, and Y. Ne'eman, Phys. Lett. **17**, 148 (1965); M. Bander and C. Itzykson, Rev. Mod. Phys. **38**, 330 (1966); Rev. Mod. Phys. **38**, 346 (1966); K. B. Wolf, Nuovo Cimento Suppl. **5**, 1041 (1967); C. P. Boyer and G. N. Fleming, Pennsylvania State University, preprint.

[4]R. Herman, *Lie Groups for Physicists* (Benjamin, New York, 1966).

[5]J. Rosen and P. Roman, J. Math. Phys. **7**, 2072 (1966); J. Rosen, Nuovo Cimento B **46**, 1 (1966); J. Math. Phys. **9**, 1305 (1968); A. Chakrabarti, J. Math. Phys. **9**, 2087 (1968).

[6]J. G. Nagel, Ann. Inst. Henri Poincaré **13**, 1 (1970).

[7]J. G. Nagel and K. T. Shah, J. Math. Phys. **11**, 1483 (1970); J. G. Nagel, J. Math. Phys. **11**, 1779 (1970).

[8]K. B. Wolf, J. Math. Phys. **12**, 197 (1971); C. P. Boyer, J. Math. Phys. **12**, 1599 (1971); C. P. Boyer and F. Ardalan, J. Math. Phys. **12**, 2070 (1971); K. B. Wolf, J. Math. Phys. **13**, 1634 (1972).

[9]R. Gilmore, J. Math. Phys. **13**, 883 (1972).

[10]V. Bargmann, Ann. Math. **48**, 568 (1947).

[11]I. M. Gel'fand and M. A. Naimark, *Unitäre Darstellungen der Klassischen Gruppen* (Akademie Verlag, Berlin, 1957); I. M. Gel'fand, M. I. Graev, and N. Ya. Vilenkin, *Generalized Functions* (Academic, New York, 1966), Vol. 5.

[12]E. Inönü and E. P. Wigner, Proc. Natl. Acad. Sci. USA **39**, 510 (1953); E. J. Saletan, J. Math. Phys. **2**, 1 (1961).

[13]These appeared in the mathematical literature with M. Gerstenhaber, Ann. Math. **79**, 59 (1964); R. Hermann, Commun. Math. Phys. **2**, 251 (1966); Commun. Math. Phys. 3, 53, 75 (1966); Commun. Math. Phys. **5**, 131 (1967); Commun. Math. Phys. **6**, 157 (1967); Commun. Math. Phys. **6**, 205 (1967); M. Lévy-Nahas J. Math. Phys. **8**, 1211 (1967); M. Lévy-Nahas and R. Seneor, Commun. Math. Phys. **9**, 242 (1968).

[14]For a coherent discussion of the classical Lie algebras see M. Gourdin, *Unitary Symmetries* (North-Holland, Amsterdam, 1967), Chaps. 12–15.

[15]We take the liberty of using the same notation for the abstract elements of the Lie algebra and their representations.

[16]R. L. Anderson and K. B. Wolf, J. Math. Phys. **11**, 3176 (1970).

[17]Since $SL(n, \mathbb{C})$ is a direct sum of two simple real Lie algebras, it will have two second-order Casimir invariants.

[18]This is the *compact* symplectic group (also called unitary-symplectic group), to be distinguished from the noncompact *real* symplectic group of transformations which preserve only the bilinear antisymmetric form.

[19]For further details on quaternions and their connection with symplectic geometry, see C. Chevalley, *Theory of Lie Groups* (Princeton U. P., Princeton, 1946), Chap. 1. For a discussion of the representations of $Sp(n)$ see P. Pajas and R. Rączka, J. Math. Phys. **9**, 1188 (1968).

[20]S. Ström, Ark. Fys. **33**, 465 (1966); Ann. Inst. Henri Poincaré A **13**, 77 (1970).

[21]C. P. Boyer and K. B. Wolf, work in progress.

[22]E. M. Stein, in *High Energy Physics and Elementary Particles* edited by A. Salam (IAEA, Vienna, 1965); G. W. Mackey, *Induced Represenations of Groups and Quantum Mechanics* (Benjamin, New York, 1967).

[23]I. M. Gel'fand and M. I. Graev, Am. Math. Soc. Transl. **2** (2), 147 (1956); G. Rosen, J. Math. Phys. **7**, 1284 (1966); I. Hulthén, Ark. Fys. **38**, 175 (1968).

[24]C. Fronsdal in *High Energy Physics and Elementary Particles,* edited by A. Salam (IAEA, Vienna, 1965); C. Fronsdal, Trieste preprint IC/66/51; W. Rühl, Nuovo Cimento **42**, 619 (1966); H. Leutwyler and V. Gorgé, Helv. Phys. Acta **41**, 171 (1968).

[25]N. Mukunda, J. Math. Phys. **10**, 897 (1969).

[26]C. P. Boyer, J. Math. Phys. **14**, 609 (1973).

Asymptotic behavior of atomic bound state wave functions*

Reinhart Ahlrichs

Institute of Physical Chemistry, University of Karlsruhe, Karlsruhe, Germany
(Received 24 October 1972; revised manuscript received 19 March 1973)

In the present paper we investigate the asymptotic properties of an exact bound state wave function φ of an n-electron atomic system within the infinite nuclear mass approximation and neglecting relativistic effects. An explicit upper bound is derived for $\|r_i^\mu \varphi\|$, where r_i denotes the distance of the ith electron from the nucleus and $\mu = 1,2,3,\cdots$. We are then able to derive upper bounds for expressions like $\|h(r_i)\varphi\|$, where $h(x)$ is an exponentially increasing function. We finally indicate an exponentially decreasing pointwise bound for φ.

I. INTRODUCTION

Two types of singularities occur in the Schrödinger eigenvalue equation of atomic and molecular systems. The first one is due to the singularities of the Coulomb potentials at points in configuration space where the coordinates of two or more particles coincide. The second one is concerned with the singular boundary condition, i.e., the case of one or more particle coordinates going to infinity.

The consequences of the Coulomb singularities on the wavefunction are well understood in the case that just two particles coincide[1,2] and lead to the well-known cusp conditions.

Concerning the asymptotic behavior of bound state eigenfunctions of Schrödinger operators, it is generally presumed that they vanish exponentially. So far, this has been rigorously proven only for one-electron molecular ions by Fox and Bazley[3] and for three-particle systems, e.g., helium like ions, by Slaggie and Wichmann.[4]

It has been proven by Schnol[5] that the eigenfunctions of an n-particle Schrödinger operator vanish exponentially in absolute value provided the potential is bounded from below and the corresponding eigenvalue is isolated and has finite multiplicity. Schmincke[6] has shown that the same result can also be obtained for certain classes of unbounded potentials, which do not include the case of Coulomb potentials, however.

In the present paper we consider bound state eigenfunctions-for the precise meaning of this term see Sec. II—of an n-electron atomic ion in the infinite nuclear mass approximation, neglecting relativistic effects. The Hamilton operator is then spin independent and we have to consider only spatial wavefunctions $\varphi(r_1,\ldots,r_n)$.

In Sec. II we first define various quantities which will be used frequently in the following derivations. We also give a brief account of some basic mathematical concepts. The asymptotic properties of a bound state eigenfunction $\varphi(r_1,\ldots,r_n)$ are then discussed in Sec. III in deriving upper bounds for expressions like $\|r_i^\mu \varphi\|$. The main results can be summarized as follows.

(a) In subsection IIIA we first derive an explicit upper bound for $\|r_i^\mu \varphi\|$, see (17), (18). It is shown that this bound cannot be improved considerably in general.

(b) In subsection IIIB it is proven that φ vanishes exponentially in the mean, i.e., $h(r_i)\varphi$ is square integrable where $h(x)$ is an exponentially increasing function. The upper bound obtained for $\|h(r_i)\varphi\|$ is also shown to be about as close as possible. We finally derive an exponentially decreasing pointwise bound for $|\varphi(r_1,\ldots,r_n)|$, which appears to be relatively poor, however.

The various explicit bounds obtained are of a simple form and depend only on the ionization potential of the state under consideration and on the nuclear charge.

II. NOTATION AND BASIC CONCEPTS

We consider solutions of the eigenvalue equation (1) in the Hilbert space $\mathcal{K} = L^2(R^{3n})$:

$$(H - E)\varphi = 0, \quad \varphi \neq 0. \tag{1}$$

For convenience φ is required to be normalized to unity:

$$\|\varphi\| = 1. \tag{2}$$

The Hamiltonian H is given by

$$H = T + V, \tag{3}$$

$$T = -\tfrac{1}{2}\sum_{i=1,n}\Delta_i \qquad V = -Z\sum_{i=1,n}(1/r_i) + \sum_{i<j}(1/r_{ij}), \tag{4}$$

where r_i denotes the coordinate vector of the ith electron with respect to the nucleus, $|r_i| = r_i, r_{ij} = r_i - r_j$, Δ is the Laplacian in three-dimensional space, Z the nuclear charge, and n the number of electrons. Whenever confusion is possible, we will write $H^{(n)}$ instead of H to indicate the dependence on n. A comment is appropriate here on the definition of H, since it is not obvious that the formal expression (3), (4) defines a self-adjoint operator on $\mathcal{K}$. It has, however, been shown by Kato[7] that a self-adjoint operator $\tilde{H} = \tilde{T} + V$ exists, which is uniquely determined by (3) and (4). $\tilde{T}$ is a generalization of T in the following sense: If $Tf, f \in \mathcal{K}$, exists locally almost everywhere and $Tf \in \mathcal{K}$, then $Tf = \tilde{T}f$.[1,7] The domains $D_{\tilde{T}}$ and $D_{\tilde{H}}$ of $\tilde{T}$ and $\tilde{H}$ coincide, which is due to the fact that each single term in V and hence V itself is relatively bounded with respect to $\tilde{T}$ with bound zero:

$$Z\sum_i \|(1/r_i)f\| + \sum_{i<j}\|(1/r_{ij})f\| \leq a\|f\| + b\|\tilde{T}f\|, \qquad f \in D_{\tilde{T}},\ a,b > 0, \tag{5}$$

where $b > 0$ can be chosen arbitrarily small.[7]

We further note that any eigenfunction φ of $\tilde{H}$ [i.e., a solution of $(\tilde{H} - E)\varphi = 0, \varphi \neq 0$], is (equivalent to) an analytical function and satisfies the differential equation (1) in any region of the configuration space R^{3n} where V is regular.[1,8] φ is still continuous at the singular points of V and has only a rather mild singularity.

In the following we have to deal with sufficiently well-behaved functions only for which $Tf = \tilde{T}f$ and hence

$Hf = \tilde{H}f$. Having this in mind, we will frequently suppress the tilde in order to keep the notation simple.

Throughout this paper we consider only bound state eigenfunctions. To give a proper definition of this term we have to consider the implications of statistics and spin. As the Hamiltonian is invariant with respect to a permutation of the particles, the eigenfunctions of H can be classified according to the irreducible representations of the symmetric group S_n. Because of the Pauli principle, only some of these are physically realizable,[9] which in the nomenclature of Wigner (Ref. 9, pp. 129–33), can be classified by the irreducible representations $\overline{D}^{(k)}, k = 0, 1, \ldots, [n/2]$. The parameter k is related to the total spin S through $S = n/2 - k$. If the eigenfunction φ under consideration belongs to $\overline{D}^{(k)}$, it can only be ionized into $(n-1)$-particle states which transform according to the irreducible representations $'\overline{D}^{(k)}$ or $'\overline{D}^{(k-1)}$ of S_{n-1}.

Let $E_{0,k}^{(n-1)}$ denote the lowest eigenvalue of $H^{(n-1)}$ belonging to $'\overline{D}^{(k)}$ and let

$$E_0^{(n-1)} = \min\left(E_{0,k}^{(n-1)}, E_{0,k-1}^{(n-1)}\right).$$

The (lowest) ionization potential ϵ of the state described by φ is then given by $\epsilon = E_0^{(n-1)} - E$.

In the following we consider only states for which

$$\epsilon = E_0^{(n-1)} - E > 0, \tag{6}$$

which are usually called bound states.

In the subsequent considerations we will use frequently the quantities η and γ defined as

$$\eta = \sqrt{2\epsilon}, \tag{7}$$

$$\gamma = Z/\eta. \tag{8}$$

It is further convenient to introduce the n-electron operator H'_i

$$H'_i = H + Z/r_i. \tag{9}$$

For the subsequent derivations we need an estimate of the following kind:

$$\langle Q_i f, (H'_i - E) Q_i f\rangle \geq \epsilon \langle Q_i f, Q_i f\rangle, \tag{10}$$

where f is a sufficiently well-behaved function which belongs to the irreducible representation $\overline{D}^{(k)}$ of S_n, and Q_i denotes an operator which depends on r_i only. We decompose $Q_i f$ with respect to the irreducible representations of the subgroup S_{n-1} (of S_n) of permutations that do not affect r_i. The only nonvanishing contributions belong to the irreducible representations $'\overline{D}^{(k-1)}$ and $'\overline{D}^{(k)}$ of S_{n-1}.[9] If we further note that

$$H'_i = H^{(n-1)}(r_1, \ldots, r_{i-1}, r_{i+1}, \ldots, r_n) - \tfrac{1}{2}\Delta_i + \sum_{j(\neq i)} \frac{1}{r_{ij}} \geq H^{(n-1)},$$

the inequality (10) follows immediately from (6) and the definition of $E_0^{(n-1)}$.

We finally define the differential operator p_{r_i}

$$p_{r_i} = \partial/\partial r_i. \tag{11}$$

It is easily shown that p_{r_i} is relatively bounded with respect to $\tilde{T}$:

$$\|p_{r_i} f\|^2 \leq -\langle f, \Delta_i f\rangle \leq \alpha \|f\|^2 + (1/\alpha)\|\tilde{T}f\|^2, \quad 0 < \alpha < \infty,$$

which then proves that $p_{r_i} f$ is well defined and square integrable if this is the case for $\tilde{T}f$.

The subsequent derivations are considerably simplified by a result due to Combes,[10] who showed that under the present conditions $r_i^\mu \varphi \in \mathfrak{K}$ and further $r_i^\mu \varphi \in D_H$ for arbitrary $\mu = 1, 2, 3, \cdots$.

III. RESULTS

A. Explicit bounds for $\|r_i^\mu \varphi\|$

The derivations presented in this subsection are essentially based on the relationship (10). We then obtain a few steps the desired bound for $\|r_i^\mu \varphi\|$; see (17).

We start from the following identity:

$$\langle r_i^\mu \varphi, (H'_i - E) r_i^\mu \varphi\rangle = Z\langle r_i^\mu \varphi, r_i^{\mu-1}\varphi\rangle - \mu\langle r_i^\mu \varphi, p_{r_i} r_i^{\mu-1}\varphi\rangle + \tfrac{1}{2}\mu(\mu-3)\|r_i^{\mu-1}\varphi\|^2, \tag{12}$$

which is easily proven if we use the eigenvalue equation in the form $(H'_i - E)\varphi = (Z/r_i)\varphi$, and apply standard differentiation techniques. It should further be noted that all terms occuring in Eq. (12) are well defined in the sense of Hilbert space theory, as has been shown in a paper by Combes[10] we have mentioned briefly at the end of Sec. II, above.

From the inequality (10) one obtains the following estimate for the lhs of Eq. (12):

$$\langle r_i^\mu \varphi, (H'_i - E) r_i^\mu \varphi\rangle \geq \epsilon \|r_i^\mu \varphi\|^2. \tag{13}$$

The term $\langle r_i^\mu \varphi, p_{r_i} r_i^{\mu-1}\varphi\rangle$, occuring on the rhs of Eq. (12), can be integrated by parts (for an almost identical derivation, see Ref. 11, p. 345):

$$\langle r_i^\mu \varphi, p_{r_i} r_i^{\mu-1}\varphi\rangle = -\tfrac{3}{2}\|r_i^{\mu-1}\varphi\|^2. \tag{14}$$

Equation (14) is valid only if the lhs of (14) is real, which is easily verified from (12). Combination of (12), (13), and (14) yields

$$\|r_i^\mu \varphi\|^2 \leq 2\gamma\eta^{-1}\|r_i^\mu \varphi\| \cdot \|r_i^{\mu-1}\varphi\| + \mu^2\eta^{-2}\|r_i^{\mu-1}\varphi\|^2, \tag{15}$$

where γ and η are defined in (7) and (8). The inequality (15) can be solved with respect to $\|r_i^\mu \varphi\|$

$$\|r_i^\mu \varphi\| \leq [\gamma + (\mu^2 + \gamma^2)^{1/2}]\eta^{-1}\|r_i^{\mu-1}\varphi\|, \quad \mu = 1, 2, \ldots. \tag{16}$$

Iterating (16), we obtain the desired explicit bound for $\|r_i^\mu \varphi\|$:

$$\|r_i^\mu \varphi\| \leq \prod_{\lambda=1,\mu} [\gamma + (\lambda^2 + \gamma^2)^{1/2}]\eta^{-\mu}\|\varphi\|, \quad \mu = 1, 2, \ldots, \tag{17}$$

where $\|\varphi\|$ can be ommitted since φ was required to be normalized to unity.

In order to make the μ dependence of (17) more transparent, we will now derive another bound which is somewhat poorer than (17), however. Using the easily verified inequality

$$\gamma + (\lambda^2 + \gamma^2)^{1/2} \leq (\lambda + \gamma)[1 + \gamma^2/2(\lambda + \gamma/2)^2],$$

one obtains

$$\prod_{\lambda=1,\mu} [\gamma + (\lambda^2 + \gamma^2)^{1/2}] \leq [\Gamma(\mu + \gamma + 1)/\Gamma(\gamma + 1)] \times \prod_{\lambda=1,\mu} [1 + \gamma^2/2(\lambda + \gamma/2)^2],$$

where $\Gamma(x)$ denotes the gamma function. The product on the rhs of the last equation is a slowly varying function of μ. If we insert the product representation of $\sinh(x)$ (see Ref. 12, p. 29), this term can be further approximated by an expression independent of μ. A simple manipulation finally yields

$$\|r_i^\mu \varphi\| \leq C(\gamma)\Gamma(\mu+\gamma+1)\eta^{-\mu}, \quad \mu = 0,1,2,\ldots, \tag{18}$$

where

$$C(\gamma) = \sqrt{2}\,\sinh(\pi\gamma/\sqrt{2})\,(\Gamma(\gamma+1)\pi\gamma \prod_{\lambda=1,\delta}(1+\gamma^2/2\lambda^2))^{-1} \tag{19}$$

and $\delta = [\gamma/2]$. The product occuring in (19) has to be set equal to unity if $\delta < 1$. The bound (18) was proven only for $\mu \geq 1$, but it is easily verified that it holds also for $\mu = 0$.

In the case of one-electron atomic ions the eigenvalue equation (1) is solvable and the bound (18) can be compared with the exact result. Let us, e.g., consider the eigenfunction $\varphi_{n,n-1}(\mathbf{r})$, where n is the principal quantum number and $l = n-1$ the angular momentum. (The magnetic quantum number m does not enter the following formulas.) The radial part $f_{n,n-1}(r)$ of $\varphi_{n,n-1}$ is

$$f_{n,n-1}(r) = N r^{n-1} e(-Zr/n),$$

where N is a normalization constant. The ionization potential ϵ is given by $\epsilon = Z^2/2n^2$, which yields $\eta = Z/n$ and $\gamma = n$. Expressing $\|r^\mu \varphi_{n,n-1}\|$ in terms of γ and η, we obtain

$$\|r^\mu \varphi_{n,n-1}\|^2 = \Gamma(2\mu+2\gamma+1)\,[\Gamma(2\gamma+1)\cdot(2\eta)^{2\mu}]^{-1},$$

which gives asymptotically for large μ

$$\|r^\mu \varphi_{n,n-1}\| \approx [\Gamma(\gamma+0.5)\Gamma(\gamma+1)]^{-1/2}(\mu+\gamma)^{-1/4} \times \Gamma(\mu+\gamma+1)\eta^{-\mu}. \tag{20}$$

Comparison of this expression with the bound (18) shows indeed that (18) has, besides the slowly varying term $(\mu+\gamma)^{-1/4}$, the correct dependence on μ and η. The factor $C(\gamma)$, however, is not too good for large γ. Considering the rather rough approximations that had to be made to derive (18), this bound for $\|r^\mu\varphi\|$ is surprisingly good.

For the following considerations it is convenient to have a bound of the kind (18) for noninteger powers of r:

$$\|r_i^{\mu+\rho}\varphi\| \leq C(\gamma)\Gamma(\mu+\gamma+1)\,(\mu+\gamma+1)^\rho \eta^{-\mu-\rho}, \tag{18'}$$

where ρ is a real number $0 \leq \rho \leq 1$. The estimate (18') is valid for $\rho = \frac{1}{2}$, since by virtue of Schwarz' inequality

$$\|r_i^{\mu+1/2}\varphi\|^2 = \langle r_i^\mu\varphi, r_i^{\mu+1}\varphi\rangle \leq \|r_i^\mu\varphi\|\cdot\|r_i^{\mu+1}\varphi\| \leq [C(\gamma)\Gamma(\mu+\gamma+1)\,(\mu+\gamma+1)^{1/2}\eta^{-\mu-1/2}]^2.$$

Repeating the same procedure we see that (18') holds for arbitrary dual fractions $\rho = j\cdot 2^{-m}$, $j = 0,1,\ldots,2^m$, and hence by continuity for any real number $0 \leq \rho \leq 1$.

B. Exponential behavior of φ

The μ and η dependence of $\|r^\mu\varphi_{n,n-1}\|$ displayed by (20) is a direct consequence of the exponentially decreasing behavior of $\varphi_{n,n-1}$. From the similarity of (18) and (20) one might expect that any bound state eigenfunction φ vanishes exponentially. This is indeed the case in a sense to be specified below. Let us beforehand define the function $h_{\beta,\nu}(x)$ for real β, integer ν, with $\nu > \beta \geq 0, x \geq 0$, through

$$h_{\beta,\nu}(x) = \sum_{\lambda=\nu,\infty}[x^{\lambda-\beta}/\Gamma(\lambda+1)], \tag{21}$$

which can also be written as

$$h_{\beta,\nu}(x) = x^{-\beta}(e^x - \sum_{\lambda=0,\nu-1}[x^\lambda/\Gamma(\lambda+1)]).$$

$h_{\beta,\nu}(x)$ behaves asymptotically like $x^{-\beta}e^x$:

$$\lim_{x\to\infty} x^\beta e^{-x} h_{\beta,\nu}(x) = 1. \tag{22}$$

We can now phrase Theorem 1.

Theorem 1: For any electronic bound state eigenfunction φ of the Hamiltonian (3), $h_{\beta,\nu}(\eta r)\varphi$ is square integrable if $\beta > \gamma + 1$, where η and γ are defined in Eqs. (7) and (8).

Proof: By virtue of (18') we have

$$\|h_{\beta,\nu}(\eta r_i)\varphi\| \leq \sum_{\lambda=\nu,\infty}\frac{\eta^{\lambda-\beta}\|r_i^{\lambda-\beta}\varphi\|}{\Gamma(\lambda+1)} \leq C(\gamma)\sum_{\lambda=\nu,\infty}\frac{\Gamma(\lambda+\gamma-[\beta])}{\Gamma(\lambda+1)}(\lambda+\gamma-[\beta])^{[\beta]+1-\beta},$$

where $[\beta]$ denotes the largest integer contained in β. To evaluate the sum on the rhs, we take advantage of

$$\Gamma(x)x^b/\Gamma(x+b) \leq 1, \quad \text{if } b \geq 1, x > 0,$$

which follows immediately from the logarithmic convexity of the gamma function; see Ref. 13, p. 4. Putting $x = \lambda+\gamma-[\beta]$ and $b = 1+[\beta]-\gamma$, we get immediately

$$\|h_{\beta,\nu}(\eta r_i)\varphi\| \leq C(\gamma)\sum_{\lambda=\nu,\infty}(\lambda+\gamma-[\beta])\eta^{-\beta}, \tag{23}$$

where the rhs does in fact converge for any $\beta > \gamma + 1$. QED

Theorem 1 cannot be improved much in general as is seen from a consideration of the particular one-electron eigenfunctions $\varphi_{n,n-1}(\mathbf{r})$ already discussed in the preceeding section. Taking into account the asymptotic behavior of $h_{\beta,\nu}(x)$ [see (22)], we easily verify that $h_{\beta,\nu}(\eta r)\varphi_{n,n-1}(\mathbf{r})$ is square integrable if $\beta > \gamma + \frac{1}{2}$, which is only slightly less restrictive than the requirement of Theorem 1: $\beta > \gamma + 1$.

We shall now prove that any bound state eigenfunction φ of H vanishes exponentially if $r_i \to \infty$. For this purpose it is convenient to write $x = (\mathbf{r}_1, \mathbf{r}_2, \ldots, \mathbf{r}_n)$. It has been shown by Kato[1] that any eigenfunction φ of the Hamiltonian (3) is Hölder continuous for $\theta < 1$, i.e., for any θ with $0 < \theta < 1$, there exists a parameter B depending on θ but independent of x and $\tilde{x}$ such that

$$|\varphi(x) - \varphi(\tilde{x})| \leq B|x-\tilde{x}|^\theta. \tag{24}$$

The inequality (24) implies

$$|\varphi(\tilde{x})|^2 \geq (|\varphi(x)| - B|x-\tilde{x}|^\theta)^2 \quad \text{for } |x-\tilde{x}| \leq R_0 = [|\varphi(x)|/B]^{1/\theta}. \tag{25}$$

We can now use (25) to verify the following chain of inequalities:

$$(\|\varphi\| + \|h_{\beta,\nu}(\eta r_i)\varphi\|^2 \geq \int_{|x-\tilde{x}|\leq R_0} |(h_{\beta,\nu}(\eta\tilde{r}_i)+1)\varphi(\tilde{x})|^2 d\tilde{x}$$

$$\geq \left(\min_{|r_i-\tilde{r}_i|\leq R_0} h_{\beta,\nu}(\eta\tilde{r}_i) + 1\right)^2 \times B^{-3n/\theta}\omega_{3n}|\varphi(x)|^{(3n/\theta+2)}\cdot[2\theta^2/3n(3n+\theta)(3n+2\theta)], \tag{26}$$

where $\omega_m = 2\pi^{m/2}/\Gamma(m/2)$ is the surface area of the unit sphere in R^m. In order to get rid of the parameter $R_0 = [|\varphi(x)|B]^{1/\theta}$ occuring in (26), we note that $|\varphi(x)|$ is bounded[1] and hence $R_0 < R_{\max}$. Since $h_{\beta,\nu}$ is essentially the exponential, see (22), there exists a constant $\tilde{K} > 0$ such that

$$\min_{|r_i - \tilde{r}_i| \leq R_0} [h_{\beta,\nu}(\eta\tilde{r}_i) + 1] \geq \min_{|r_i - \tilde{r}_i| < R_{\max}} [h_{\beta,\nu}(\eta\tilde{r}_i) + 1] \geq \tilde{K}[h_{\beta,\nu}(\eta r_i) + 1].$$

Combining the last inequality with (26), we finally obtain

$$|\varphi(x)| \leq K\,[h_{\beta,\nu}(\eta r_i) + 1]^{-\kappa}, \quad \beta > \gamma + 1, \tag{27}$$

where $\kappa = (\frac{3}{2}n/\theta + 1)^{-1}$ and K is a suitable chosen parameter independent of x. By virtue of the asymptotic behavior of $h_{\beta,\nu}$, see (22), (27), yields for sufficiently large r_i

$$|\varphi(x)| \leq K r_i^{\kappa\beta} e^{-\eta\kappa r_i}. \tag{28}$$

If we put $s = \max_i r_i$, the following inequality (28′) is an immediate consequence of (28), provided s is sufficiently large:

$$|\varphi(x)| \leq K s^{\kappa\beta} e^{-\eta\kappa s}. \tag{28′}$$

IV. DISCUSSION

In the preceding sections we have derived bounds for $\|r_i^{\mu}\varphi\|$ and $\|h_{\beta,\nu}(\eta r_i)\varphi\|$, where φ denotes as before an atomic bound state eigenfunction. The corresponding bounds (18) and (23) provide information about the asymptotic behavior of φ in the mean for the case $r_i \to \infty$. It has further been shown that (18) and (23) cannot be improved considerably in general.

One can also use the results presented above to obtain bounds for expressions involving many-electron operators. Let us, e.g., consider R defined as

$$R = \left(\sum_{i=1,n} r_i^2\right)^{1/2}.$$

It is then easily proven that

$$R^{2\mu} \leq n^{\mu-1} \sum_{i=1,n} r_i^{2\mu}$$

which in combination with (18) yields

$$\|R^{\mu}\varphi\| \leq C(\gamma)\Gamma(\mu + \gamma + 1)n^{\mu}\eta^{-\mu} \tag{29}$$

and further, in analogy to (23),

$$\|h_{\beta,\nu}(\eta R/n)\varphi\| \leq C(\gamma) \sum_{\lambda=\nu,\infty} (\lambda + \gamma - [\beta])^{\gamma-\beta}, \tag{30}$$

provided $\nu > \beta > \gamma + 1$.

The bounds (29) and (30) are rather poor, however. The relationship (30) proves that φ vanishes in the mean at least like $e^{-\eta R/n}$, whereas one would expect a decay like $e^{-\eta R}$. It has in fact recently been shown by O'Connor[14] that $\varphi e^{\eta' R} \in \mathcal{K}$ if $\eta' < \eta$. From O'Connor's analysis it follows, of course, that $\varphi e^{\eta' r_i} \in \mathcal{K}$, if $\eta' < \eta$, this result is, however, weaker than the estimate (23) which includes the case $\eta' = \eta$. We, furthermore, note that O'Connor's treatment does not furnish explicit bounds like (23) or (18) derived in the present paper.

The exponentially decreasing pointwise bound (28) is rather poor too. Partly stimulated by the present study, Simon[15] has recently shown that $|\varphi| < A(\eta')e^{-\eta' R}$, if $\eta' < \eta$. The proof of Simon is based on the treatment of O'Connor.

We finally make some comments concerning the generalization of the results obtained in this study. It could appear that the derivations given in Sec. III are restricted to the particular Hamiltonian specified in Eqs. (3) and (4). This is not the case, however. The considerations presented in this work can in fact be extended to Hamiltonians of the form $H = T + W$, where W is a so called Kato potential.[1] We then define H_i' as

$$H_i' = H + \tau/|r_i - r_{i,0}|,$$

where the parameters τ and $r_{i,0}$ have to be chosen such that the relationship (31) holds [which is analogous to (10) above],

$$\langle Q_i f, (H_i' - E)Q_i f\rangle \geq \epsilon(\tau)\,\langle Q_i f, Q_i f\rangle, \tag{31}$$

for $\epsilon(\tau) > 0$. The present method breaks down, however, if (31) cannot be fulfilled. Under the presupposition that (31) holds, all bounds derived above are obviously valid if we replace r_i by $|r_i - r_{i,0}|$, Z by τ and ϵ by $\epsilon(\tau)$. The parameters τ and $r_{i,0}$ can then be chosen to optimize the various bounds. In the atomic case the optimal choice is, of course, $r_{i,0} = 0$ and $\tau = Z$, which in turn has the advantage that ϵ can be interpreted as the ionization potential.

ACKNOWLEDGMENTS

The author is indebted to Dr. L. Thomas for valuable comments concerning the manuscript and to Dr. W. Kutzelnigg and Dr. V. Staemmler for numerous discussions. I further thank Professor B. Simon for bringing Ref. 10 to my attention and Dr. A. J. O'Connor for allowing me to see his work before publication.

*A preliminary note on the pertinent results of this study has been published in Chem. Phys. Lett. **15**, 609 (1972).
[1]T. Kato, Commun. Pure Appl. Math. **10**, 151 (1957).
[2]W. A. Bingel, Z. Naturforsch. A **18**, 1249 (1963).
[3]N. W. Bazley and D. W. Fox, Int. J. Quantum Chem. **3**, 581 (1969).
[4]E. L. Slaggie and E. H. Wichmann, J. Math. Phys. **3**, 946 (1962).
[5]E. E. Schnol, Mat. Sbornik (N.S.) 42 (84), 273 (1957) (in Russian); see also I. M. Glazman, *Direct methods of qualitative spectra analysis of singular differential operators* (Moscow, 1963) (English translation, Israel Program for Scientific Translations, Jerusalem, 1965).
[6]U.-W. Schmincke, Math. Z. **111**, 267 (1969).
[7]T. Kato, Trans. Am. Math. Soc. **70**, 195 (1951).
[8]E. Hopf, Math. Z. **34**, 194 (1932).
[9]E. P. Wigner, *Group theory and its application to the quantum mechanics of atomic spectra* (Academic, New York, 1959).
[10]J. M. Combes, Nuovo Cimento A **64**, 111 (1969).
[11]T. Kato, *Perturbation theory for linear operators* (Springer, Berlin, 1966).
[12]K. Knopp, *Funktionentheorie* (de Gruyter, Berlin, 1955), Vol. II.
[13]Jahnke-Emde-Lösch, *Tables of higher functions* (Teubner, Stuttgart, 1960).
[14]A. J. O'Connor, to be published.
[15]B. Simon, private communication.

Energy eigenvalues of a bounded centrally located harmonic oscillator

Richard Vawter

Department of Physics, Western Washington State College, Bellingham, Washington 98225
(Received 30 March 1973)

In the study of the effects of finite boundaries on the magnetic properties of a solid, one encounters the problem of finding the energy eigenvalues of a one-dimensional harmonic oscillator located in a potential enclosure. Series expansion techniques are applied to solve this problem for a harmonic oscillator located at the center of an infinitely high potential well. An analytical expression for the energy eigenvalues is found as a function of the size potential enclosure L, the quantum state n, oscillator frequency ω, and the mass of the particle m. The first order approximation of this expression is given by $E = E_{osc}\coth(E_{osc}/E_{box})$ where $E_{osc} = \hbar\ \omega(n + 1/2)$ is the energy eigenvalues of an unbound harmonic oscillator and $E_{box} = (2m/\hbar^2)\ (n + 1)^2\pi^2/L^2$ is the energy eigenvalues of a free particle in an infinitely high well. For the ground state this approximation is better than 1% for all values of L, ω, and m.

I. INTRODUCTION

The problem of finding the energy eigenvalues of a constrained harmonic oscillator is of interest in several areas of physics. This mainly comes about because the squared or harmonic term in a Taylor's series expansion of the potential energy gives the most important contribution to the Hamiltonian. Our interest in this problem arises from a study of the effects of finite boundaries on the magnetic properties of a metallic solid. Using the one-electron approximation, the problem becomes that of finding the energy eigenvalues of a single electron in a magnetic field which is also in a potential well or box. This problem can further be reduced to that of a one-dimensional harmonic oscillator in a box. The center of the oscillator can be anywhere inside or outside the box. In a previous paper the energy eigenvalues for this problem were found both by numerical extraction of the eigenvalues from the exact series solution of the bound-oscillator differential equation and also by applying the WKB method to the bound-oscillator differential equation.[1] The WKB solution, although generally quite good, gives its poorest agreement with numerical solution (10% error) for the energy eigenvalues in the region of most interest, namely the eigenvalues corresponding to those energy eigenstates for which the classical turning point of the unconstrained oscillator is near the edge of the potential well. In this paper by restricting the center of the oscillator to the simpler case of being at the center of the potential enclosure we are able to obtain an analytical expression for the energy eigenvalues in this region. Hopefully the methods of this paper can be expanded at a later date to include those cases where the center of the oscillator is not at the center of the potential well.

The Hamiltonian of a harmonic oscillator in the center of an infinitely high potential well of length L can be written as

$$H = (1/2m)p^2 + \tfrac{1}{2}m\omega^2x^2 + V(x),$$

where

$$V(x) = \begin{cases} 0, & -L/2 < x < L/2, \\ \infty, & |x| \geq L/2. \end{cases}$$

The problem of finding the energy eigenvalues as a function of the length of the box L, the frequency (or strength) of the oscillator ω, the mass m, and the quantum state n can be reduced to solving the dimensionless Sturm–Lionville differential equation

$$\psi''(\tau) + (\lambda - \tau^2)\psi(\tau) = 0 \tag{1}$$

with the boundary conditions that the wavefunction $\psi(\tau)$ vanish at the walls, $\psi(\pm a) = 0$. Here τ, λ, l, and a are dimensionless parameters defined by

$$\tau = (m\omega/\hbar)^{1/2}x, \quad \lambda = 2E/\hbar\omega,$$
$$a = l/2, \quad l = (m\omega/\hbar)^{1/2}L.$$

There are two important limiting cases for the eigenvalues of this problem. When an energy level λ lies well down inside the harmonic portion of the potential well (region I in Fig. 1), the eigenvalues approach those of an unbound harmonic oscillator, $\lambda \to 2n + 1$ for $\lambda \ll l^2/4$. For an unbound oscillator the eigenfunctions of (1) are given by

$$\psi_n(\tau) = e^{-\tau^2/2}H_n(\tau),$$

where $H_n(\tau)$ are Hermite polynominals. Since $\psi_n(\tau)$ falls off as $e^{-\tau^2/2}$ beyond the classical turning point, the free oscillator wavefunction will be very close to zero at the walls of the potential enclosure provided the turning points are not near the walls. For an unbound oscillator the point given by $\lambda = \tau^2$ represents the turning point of the oscillator. Consequently, τ will be close to $2n + 1$ provided $2n + 1$ is not close to $l^2/4 = a^2$.

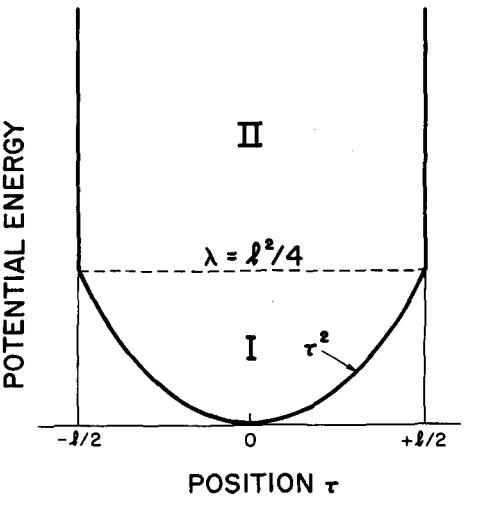


FIG. 1 Potential energy of a dimensionless harmonic oscillator centrally located in an infinitely high potential well. Region I is the area bounded by $\lambda = l^2/4$ and τ^2.

At the other extreme when an energy level is very high above the harmonic portion of the potential well (region II in Fig. 1) the eigenvalues approach those of a free particle in a box, $\lambda \to (n+1)^2\pi^2/l^2$ for $\lambda \gg l^2/4$ where $n = 0, 1, 2, \cdots$. This is not surprising since classically the oscillator will bounce off the walls long before it can reach its normal turning point so that it sees little effect of the harmonic portion of the potential well. It is the transition region around $\lambda = l^2/4$ between the unbound oscillator states and the particle in a box states that is of the most interest.

The centrally located bound harmonic oscillator has been investigated by several authors.[1-5] Their results for the energy eigenstates can be divided into two groups. In the first group the eigenvalues are given as a table of numbers.[1,5] Here the eigenvalues have been extracted numerically from the exact solution of (1). Although the eigenvalues can be found for most of the ranges of the parameters involved, the results are expressed as numbers as opposed to an analytical expression. In the second group asymptotic expressions for the eigenstates are found in the limiting ranges discussed above.[1-4] We have overcome these limitations by showing that all energy eigenvalues are given by a series expansion of the form

$$\lambda = \lambda_0 \coth[(\lambda_0/\lambda_b) + a_1(\lambda_0/\lambda_b)^3 + a_2(\lambda_0/\lambda_b)^5 + \cdots], \tag{2}$$

where λ_0 is the dimensionless free oscillator eigenstates

$$\lambda_0 = 2n + 1$$

and λ_b is the dimensionless free particle in a box eigenstates

$$\lambda_b = (n+1)^2\pi^2/l^2.$$

Here the coefficients a_k in this series depend only on the quantum number n.

In the rest of this paper we (1) display a method for determining the successive coefficients a_k and (2) discuss the error made in truncating the series.

II. SERIES SOLUTION OF EIGENVALUES

To match the limiting forms of λ, we choose the following series as a representation of λ as a function of the quantum state n and the size of the dimensionless box $l = 2a$:

$$\lambda = (2n+1)\coth F(a^2), \tag{3}$$

where

$$F(x) = \sum_{k=0}^{\infty} c_k x^{2k+1}. \tag{4}$$

We have chosen the above form for λ because of the asymptotic behavior of the hyperbolic cotangent function;

$$\coth z \to 1/z \text{ as } z \to 0, \tag{5a}$$

$$\coth z \to 1 \text{ as } z \to \infty. \tag{5b}$$

Recalling that $\lambda \to 2n+1$ when the size of the potential box l becomes large, the form (3) for λ possess the proper behavior because of (5b). The other limiting behavior, $l \to (n+1)^2\pi^2/l^2$ when l is sufficiently small, can be used to determine the first coefficient c_0 of $F(a^2)$. By keeping only the first term of $F(a)^2$ when l is small, the asymptotic limit (5a) implies that (3) approaches $(2n+1)c_0a^2$. Therefore,

$$c_0 = (2n+1)/(n+1)^2 4\pi^2$$

if (3) is to approach $(n+1)^2\pi^2/l^2$ as l becomes small.

Thus the expression

$$\lambda = (2n+1)\coth[(2n+1)l^2/(n+1)^2\pi^2]$$

is a first order approximation of λ which has the proper asymptotic limits. Amazingly this approximation gives very accurate results over a large range of parameter l or n (see Sec. IV).

To obtain higher order approximation, we need to determine the successive coefficients c_k of $F(a^2)$ in (4). To accomplish this, we will use the familiar series solution of the differential equation (1) which leads to Hermite polynomials in the unbound case. The boundary conditions that the wavefunctions vanish at the boundaries will give us a restriction that will be used to determine the coefficients c_k.

Let

$$\psi(\tau) = e^{-\tau^2/2}U(\tau)$$

and

$$U(\tau) = \sum_{k=0}^{\infty} a_k \tau^k.$$

Substitution of the above in (1) gives the even and odd solutions

$$U_1(\tau) = a_0\left(1 + \sum_{k=1}^{\infty} (1-\lambda)(5-\lambda)\cdots(4k-3-\lambda)\tau^{2k}/(2k)!\right), \tag{6a}$$

$$U_2(\tau) = a_1\left(\tau + \sum_{k=1}^{\infty} (3-\lambda)(7-\lambda)\cdots(4k-1-\lambda)\tau^{2k+1}/(2k+1)!\right). \tag{6b}$$

The central location of the oscillator in the potential well means that the two solutions are independent. The constants a_0 and a_1 are determined by normalization of $U_1(\tau)$ and $U_2(\tau)$. The boundary conditions become $U_1(a) = 0$ and $U_2(a) = 0$.

By expanding $U_1(a)$ and $U_2(a)$ in a power series in a, the boundary conditions imply that the coefficient of each term in the power series in a must vanish since the value of a can be arbitrarily fixed.

To expand $U_1(a)$ and $U_2(a)$ in a power series in a, we must also expand λ in a power series in a, and also the product $(1-\lambda)(5-\lambda)\cdots(4k-3-\lambda)$ or $(3-\lambda)(7-\lambda)\cdots(4k-1-\lambda)$. To this end, we write

$$\lambda = \sum_{k=0}^{\infty} b_k a^{4k-2}. \tag{7}$$

We have chosen this form for convenience of presentation in this paper. We know that had we chosen a power series with every power in a, the boundary conditions would have eventually shown that only the powers used in (7) would be different from zero.

The c_k can be related to the b_k by expanding λ as given by (3) and (4) in a power series in a and comparing coefficients with λ given by (7). To do this, we note that

$$\coth z = \sum_{k=0}^{\infty} 2^{2k}B_{2k}z^{2k-1}/(2k)!, \qquad |z| < \pi,$$

where B_{2k} are $2k^{\text{th}}$ Bernoulli numbers.[6] Therefore,

$$\coth F(a^2) = \sum_{k=0}^{\infty} \frac{2^{2k}B_{2k}F^{2k-1}(a^2)}{(2k)!}. \tag{8}$$

To expand F^{2k-1} as a power series in a, let us look at the k^{th} power of $F(x)$. Since the terms in the expansion of $F(x)$ go as x^{4n+2}, it follows that $F^k(x)$ goes as x^{4n+2k}, so let us define C_n^k to be the coefficients of the expansion of F^k,

$$F^k(x) = \sum_{n=0}^{\infty} C_n^k x^{4n+2k}. \tag{9}$$

Note that the symbol "n" will sometimes be used as a summation index for convenience and should not be confused with the quantum number n. Appendix A gives the relationship between C_n^k and c_n. Using (9) in (8), we have

$$\lambda = (2n+1)\sum_{k=0}^{\infty}\sum_{l=0}^{\infty} 2^{2k}B_{2k}C_l^{2k-1}a^{4l+4k-2}/(2k)!$$

and a comparison with (7) shows that

$$b_k = (2n+1)\sum_{t=0}^{k}\frac{2^{2t}B_{2t}C_k^{2t-1}}{(2t)!}. \tag{10}$$

With the aid of Appendix A the relationship between the first few terms are

$$c_0 = (2n+1)/b_0, \tag{11a}$$

$$c_1 = {c_0}^3/3 - b_1{c_0}^2/(2n+1), \tag{11b}$$

$$c_2 = -{c_0}^5/45 - b_2{c_0}^2/(2n+1) + b_1{c_0}^3/3 + {b_1}^2/c_0. \tag{11c}$$

As noted earlier, we will be able to determine c_0, c_1, and c_2 once b_0, b_1, and b_2 are known.

Returning to the problem of using $U_1(a) = 0$ to determine b_k, we need to expand $(1-\lambda)(5-\lambda)\cdots(4n-3-\lambda)$ in a power series in a. Let A_k^n be defined by

$$(1-\lambda)(5-\lambda)\cdots(4n-3-\lambda) = \sum_{k=0}^{n} A_k^n\lambda^k \tag{12}$$

for $n = 1, 2, 3, \cdots$. Appendix B shows how the value of coefficients A_k^n can be determined. Expanding λ^k in a power series in a, let B_k^n be defined by

$$\lambda^k = \sum_{n=0}^{\infty} B_n^k a^{4n-2k}. \tag{13}$$

Since λ is also given by (7) the coefficients b_n and B_n^k are related. Appendix A shows how the B_n^k can be determined from a knowledge of b_n. Using (12) and (13) in $U_1(a) = 0$, we have

$$1 + \sum_{n=1}^{\infty}\sum_{k=0}^{n}\sum_{l=0}^{\infty}\frac{A_k^n B_l^k a^{4l-2k+2n}}{(2n)!} = 0 \tag{14}$$

Examination of the powers of a show that the powers of a go as $4m$ or $4m+2$ so that we can rewrite the last expression as

$$\sum_{m=0}^{\infty}\mathfrak{D}_{4m}a^{4m} + \sum_{m=0}^{\infty}\mathfrak{D}_{4m+2}a^{4m+2} = 0,$$

where

$$\mathfrak{D}_0 = 1 + \sum_{n=1}^{\infty} A_n^n B_0^n/(2n)!, \tag{15a}$$

$$\mathfrak{D}_{4m} = \sum_{n=1}^{\infty}\frac{A_n^n B_m^n}{(2n)!} + \sum_{l=0}^{m-1}\sum_{n=2m-2l}^{\infty}\frac{A_{n-2m+2l}^n B_l^{n-2m+2l}}{(2n)!} \tag{15b}$$

for $m \geq 1$, and

$$\mathfrak{D}_{4m+2} = \sum_{l=0}^{m}\sum_{n=2m-2l+1}^{\infty}\frac{A_{n-2m+2l-1}^n B_l^{n-2m+2l-1}}{(2n)!} \tag{16}$$

Since a can have any value we choose, $\mathfrak{D}_{4m}$ and $\mathfrak{D}_{4m+2}$ must both be equal to zero if (14) is to vanish. These equations determine the coefficients b_n. For example, Appendices A and B show that $A_n^n = (-1)^n$ and $B_0^n = {b_0}^n$ so that

$$\mathfrak{D}_0 = 1 + \sum_{n=1}^{\infty}\frac{(-1)^n{b_0}^n}{(2n)!} = \cos\sqrt{b_0}$$

Thus $\mathfrak{D}_0 = 0$ implies that

$$b_0 = (2s+1)^2\pi^2/4 \tag{17}$$

for $s = 0, 1, 2, \cdots$. It turns out that either $\mathfrak{D}_{4m}$ or $\mathfrak{D}_{4m+2}$ will determine the coefficients b_m as a recurrence relation in lower order coefficients. Therefore, we need only look at $\mathfrak{D}_{4m} = 0$ to determine the sucessive coefficients b_m.

Appendix B shows that A_{n-i}^n can be written as

$$A_{n-i}^n = (-1)^{n-i}\sum_{j=i}^{2i}\binom{n}{j}\mathfrak{A}_j^i, \tag{18}$$

where the values of $\mathfrak{A}_j^i$ can be determined from the recurrence relation (B5) in Appendix B and $\binom{n}{j}$ is the binomial coefficient. Appendix A shows that B_k^n can be written as

$$B_k^n = \sum_{l=0}^{k}\binom{n}{j}{b_0}^{n-l}\mathfrak{B}_l^k, \tag{19}$$

where $\mathfrak{B}_l^k$ are functions of $b_1, b_2, b_3, \ldots$ but not b_0. Note that $\mathfrak{B}_l^k$ can be obtained using the recurrence relation (A3) in Appendix A.

To evaluate $\mathfrak{D}_{4m}$ and $\mathfrak{D}_{4m+2}$, we observe from (15) and (16) that they contain terms of the form

$$\sum_{n=i}^{\infty}\frac{A_{n-i}^n B_l^{n-i}}{(2n)!} = \sum_{n=i}^{\infty}\frac{(-1)^{n-i}}{(2n)!}\sum_{j=i}^{2i}\binom{n}{j}\mathfrak{A}_j^i\sum_{s=0}^{l}\binom{n-i}{s}{b_0}^{n-i-s}\mathfrak{B}_s^l$$

$$= \sum_{s=0}^{l}\frac{\mathfrak{B}_s^l}{s!}\frac{d^s}{db_0^s}\left[\sum_{j=i}^{2i}\frac{(-1)^i\mathfrak{A}_j^i}{j!}b_0^{j-i}\frac{d^j}{db_0^j}\left(\sum_{n=i}^{\infty}\frac{(-1)^n{b_0}^n}{(2n)!}\right)\right].$$

Since $j \geq i$, the jth derivative of the last sum over n starting at $n = i$ is equal to jth derivative of the sum starting at $n = 1$, which is $\cos\sqrt{(b_0)}$. If we define

$$Z_i^s(x) = \frac{d^s}{dx^s}\left(\sum_{j=i}^{2i}\frac{(-1)^i\mathfrak{A}_j^i}{j!}x^{j-i}\frac{d^j}{dx^j}(\cos\sqrt{x})\right),$$

then

$$\sum_{n=i}^{\infty}\frac{A_{n-i}^n B_l^{n-i}}{(2n)!} = \sum_{s=0}^{l}\frac{\mathfrak{B}_s^l Z_i^s(b_0)}{s!}. \tag{21}$$

Using this relationship, we find for $m \geq 1$ that

$$\mathfrak{D}_{4m} = \sum_{l=0}^{m}\sum_{s=0}^{l}\frac{\mathfrak{B}_s^l Z_{2m-2l}^s}{s!}. \tag{22}$$

In a similar manner

$$\mathfrak{D}_{4m+2} = \sum_{l=0}^{m}\sum_{s=0}^{l}\frac{\mathfrak{B}_s^l Z_{2m-2l+1}^s}{s!}. \tag{23}$$

Since $\mathfrak{B}_1^m = b_m$, we can expand (22) so that $\mathfrak{D}_{4m} = 0$ gives for $m \geq 1$

$$b_m = \left(Z_{2m}^0 + \sum_{l=1}^{m-1} b_l Z_{2m-2l}^1 + \sum_{l=2}^{m}\sum_{s=2}^{l}\frac{\mathfrak{B}_s^l Z_{2m-2l}^s}{s!}\right)\Big/(-Z_0^1), \tag{24}$$

where Z_i^s are understood to be evaluated at b_0. This expression is a recurrence relation for b_m in terms of $b_0, b_1, \cdots, b_{m-1}$, which we will use to determine the coefficient b_m.

For b_1, Appendix C shows that Z_k^n can be reduced to

$$Z_k^n(x) = \cos\sqrt{x} \sum_{l=1} \frac{(-1)^l H_{2l}^{nk}}{x^{k+n-l}} + \frac{\sin\sqrt{x}}{\sqrt{x}} \sum_{l=1} \frac{(-1)^l H_{2l-1}^{nk}}{x^{k+n-l}}, \tag{25}$$

where H_s^{nk} are numbers that can be determined from (C8) in Appendix C. Setting $m = 1$ in (24), we find

$$b_1 = Z_2^0/(-Z_0^1) = H_1^{02}/b_0 - H_3^{02}.$$

Note that only the second term in (25) survives when evaluated at b_0 since $\mathfrak{D}_0 = 0$ implies that $\cos\sqrt{b_0} = 0$.

Using Table IV (Appendix C), we find

$$b_1 = \tfrac{1}{3} - 1/2b_0. \tag{26}$$

Similarly

$$b_2 = 1/45b_0 - 5/12{b_0}^2 + 7/8{b_0}^3. \tag{27}$$

Turning now to the odd solution $U_2(\tau)$, we can again follow the same procedure used on $U_1(\tau)$. Except for b_0, the boundary condition $U_2(a) = 0$ gives the same relationship between the coefficient b_n and b_0 as does (26) or (27) although not necessarily in the same form as (24). The first term in the expansion of $U_2(a)$ in powers of a analogous to (15a) gives $(\sin\sqrt{b_0})/\sqrt{b_0} = 0$. Thus for the odd solution $b_0 = (s+1)^2\pi^2$, where $s = 0, 1, 2, \cdots$. We can combine the even and odd solutions together as

$$b_0 = (n+1)^2(\pi^2)^2, \quad n = 0, 1, 2, \cdots. \tag{28}$$

Having determined b_k we can determine c_k using (10). For the first three terms we find

$$c_0 = (2n+1)4/(n+1)^2\pi^2, \tag{29a}$$

$$c_1 = {c_0}^3[\tfrac{1}{3} + \tfrac{1}{2}(2n+1)^{-2}] - {c_0}^2[\tfrac{1}{3}(2n+1)^{-1}], \tag{29b}$$

$$\begin{aligned} c_2 = {} & {c_0}^5[\tfrac{1}{5} + \tfrac{1}{2}(2n+1)^{-2} - \tfrac{5}{8}(2n+1)^{-4}] \\ & + {c_0}^4[\tfrac{1}{12}(2n+1)^{-3} - \tfrac{1}{3}(2n+1)^{-1}] \\ & + {c_0}^3[\tfrac{4}{45}(2n+1)^{-2}]. \end{aligned} \tag{29c}$$

If higher order coefficients are needed, they can best be found by computerizing the above procedure. This is necessary because of the complexities involved in finding the value of such terms as H_s^{nk} for large n, k, and s. Since all the expressions used are in the form of recurrence relationship, the above procedure lends itself nicely to computerization.

III. OSCILLATOR LOCATED AT ONE EDGE

If we place one of the walls of our potential enclosure at $\tau = 0$, then this is equivalent to looking at a bound oscillator whose center is at one edge of the potential enclosure. The wavefunctions (6) are still valid solutions, and $U(0) = 0$ implies $a_0 = 0$. Only the odd solution survives. When it is evaluated at a, $U_2(a) = 0$, we obtain the same condition as before in Sec. II. As a result the odd solution we obtained for the centrally located oscillator will be valid for an oscillator located at one edge if l is replaced by $2l$ and n is replaced by $2n+1$. The first order approximation to λ becomes

$$\lambda = (4n+3)\coth(4n+3)l^2/(n+1)^2\pi^2.$$

IV. ERROR ANALYSIS

Using (29) a more convenient form for equations (3) and (4) of the solution of λ is given by

$$\lambda = (2n+1)\coth[z^2 + a_1 z^6 + a_2 z^{10} + \cdots], \tag{30}$$

where

$$z^2 = \lambda_0/\lambda_b = (2n+1)l^2/(n+1)^2\pi^2, \tag{31a}$$

$$a_1 = \tfrac{1}{3} + \tfrac{1}{2}(2n+1)^{-2} + \tfrac{1}{12}(n+1)^2\pi^2(2n+1)^{-2}, \tag{31b}$$

$$\begin{aligned} a_2 = {} & \tfrac{1}{5} + \tfrac{1}{2}(2n+1)^{-2} - \tfrac{1}{12}(n+1)^2\pi^2(2n+1)^{-2} \\ & - \tfrac{5}{8}(2n+1)^{-4} + \tfrac{1}{48}(n+1)^2\pi^2(2n+1)^{-4} \\ & + \tfrac{1}{180}(n+1)^4\pi^4(2n+1)^{-4}. \end{aligned} \tag{31c}$$

If $\lambda_0 < \lambda_b$ (i.e., z is less then one) we expect the first few terms in (30) to give good agreement with the exact eigenvalues since the maximum value of a_1 is 0.1277 and a_2 is 0.0282 (which occur only when n becomes large). As a function of the quantum state n, z has its maximum value at $n = 0$ so that $z < 1$ for all n if the ground state of the unbound oscillator is lower than the ground state of the free particle in a box:

$$1 < \pi^2/l^2 \quad \text{or} \quad \tfrac{1}{2}\hbar\omega < (\hbar^2/2m)\pi^2/L^2.$$

An equivalent statement is that $z < 1$ for all n if the dimensionless free particle in a box's ground state eigenvalue is greater than one, $\pi^2/l^2 > 1$. When the dimensionless ground state eigenvalue is smaller than one, z will not be less than one for all quantum states. If N is the minimum value for n for which $z \leq 1$, then for any $n > N$ we will have $z < 1$. The quantum state N can be determined from the condition $z \leq 1$ by solving $(2n+1)l^2/(n+1)^2\pi^2 \leq 1$ for N

$$N \geq (l^2/\pi^2 - 1) + l^2/\pi^2(l^2/\pi^2 - 1)^{1/2}.$$

Table I is a comparison for the ground state between the exact energy eigenvalues and the truncated expression retaining only the first term, $\lambda = (2n+1)\coth z^2$. Here the exact eigenvalues were extracted numerically using a method described in an earlier paper.[1] If the second term in (30) is included, the results agree with the computer extracted results to within the limit of accuracy of the computer values, which is one part in 10^5. A direct comparison between (30) and the numerical solutions is not possible for large l and n because of the limitations of the computer program used to find the exact eigenvalues.

When z is greater than one, we expect the higher order terms of (30) to contribute significantly. In fact, the series may even diverge if z is sufficiently large. One saving factor is that the series appears in the argument of a hyperbolic cotangent function which is close to one even if its argument is no larger than 4 ($\coth 4 = 1.000\,671$). This means that before z can become too large, (30) is already very close to the unbound oscillator states $2n+1$. Therefore, one may not really be interested in finding λ from (30) for large values of z since under these circumstances λ will be very close to $(2n+1)$. For example, in Table I we see that even when $z \geq 1$ the first order approximation still gives agreement better than 1% for the ground state. Moreover, as z becomes large, the percent error decreases as λ gets closer to the unbound ground state value of one.

Another way to look at the case when $z \geq 1$ is to note that truncating (30) gives a value for λ which is too

TABLE I. Comparison of the exact energy eigenvalues found numerically and the first order approximation as a function of the size of the potential well l for the ground state $n = 0$.

l	Exact λ	$\lambda = (2n+1)\coth z^2$	% Error	z
1	9.90225	9.90335	0.0110	0.31830
1.5	4.45979	4.46221	0.0543	0.47746
2.0	2.59691	2.60103	0.1586	0.63661
2.5	1.77893	1.78478	0.3290	0.79570
3.0	1.37786	1.38496	0.5156	0.95492
3.5	1.17497	1.18232	0.6255	1.14408
4.0	1.07492	1.08132	0.5958	1.27323
4.5	1.02893	1.03358	0.4514	1.43239
5.0	1.00990	1.01269	0.2758	1.59154
5.5	1.00297	1.00436	0.1381	1.75070
6.0	1.00076	1.00138	0.0594	1.90985

large, so that the truncated approximation of λ will be an upper bound for the exact value of λ (for example, see Table I). Since the exact value of λ is always greater than $2n + 1$, we have that

$$2n + 1 < \lambda < (2n + 1)\coth z^2.$$

From this we can conclude (as before) that when z is large the upper bound of λ is very close to $2n + 1$ so that λ must be very close to $2n + 1$.

V. SUMMARY

We have shown that (3) or (30) is an exact expression for the energy eigenvalues of a bound harmonic oscillator whose center is at the center of the potential well. Moreover, if l is replaced by $2l$ and n is replaced by $2n + 1$ in (30), the resulting expression gives the energy eigenvalues when the center of the oscillator is located at one edge of the potential enclosure.

When the natural frequency of the oscillator ω, the mass of the particle m, and the size of the well L are such that

$$\tfrac{1}{2}m\omega \ll (\hbar^2/2m)\pi^2/L^2,$$

one can truncate (30) to give an approximate expression for the eigenvalues since z will be less than one in this case for all quantum states. The truncated solution may also be used for any quantum state where the corresponding free oscillator state is less than the corresponding free particle in a box state, which occurs when the quantum number

$$n \geq (l^2/\pi^2 - 1) + (l^2/\pi^2)(l^2/\pi^2 - 1)^{1/2}.$$

The truncated solution of λ will be an upper bound of the exact value of λ. Since $2n + 1$ is always a lower bound of λ, the truncated expression can be used to obtain an approximate value of λ even in those cases where $z > 1$. For the ground state the truncated approximation (retaining only the first term) gives an expression which is always better than 1% for all ranges of the values of the parameters m, ω, and L.

APPENDIX A: THE RAISING OF A POWER SERIES TO A POWER

Let X be the power series

$$X = \sum_{n=0}^{\infty} b_n x^{4n}.$$

Define B_n^k to be the coefficient of the power series of the k^{th} power of X,

$$X^k = \sum_{n=0}^{\infty} B_n^k x^{4n}.$$

Since $X^1 = X$ and $X^0 = 1$, we have

$$B_n^1 = b_n, \quad B_0^0 = 1,$$
$$B_n^0 = 0 \quad \text{for } n \geq 1.$$

Using $X^k = X^{k-1}X$ and equating the coefficients of powers of x, we obtain the following recurrence relation for $n \geq 0$:

$$B_n^k = \sum_{l=0}^{n} b_{n-l} B_l^{k-1}. \tag{A1}$$

For the first few coefficients of X^k this recurrence relation yields

$$B_0^k = b_0{}^k,$$
$$B_1^k = k b_0{}^{k-1} b_1,$$
$$B_2^k = k b_0{}^{k-1} b_2 + k(k-1) b_0{}^{k-2} b_1{}^2/2!,$$
$$B_3^k = k b_0{}^{k-1} b_3 + k(k-1) b_0{}^{k-2} b_1 b_2 + k(k-1)(k-2) b_0{}^{k-3} b_1{}^3/3!.$$

The form of the above terms indicates that we can write B_k^n in the following form

$$B_n^k = \sum_{l=0}^{n} \binom{k}{l} b_0{}^{k-l} \mathfrak{B}_l^n, \tag{A2}$$

where $\mathfrak{B}_l^n$ does not depend on b_0. Here $\binom{k}{l}$ is the binomial coefficient.[6] To obtain a recurrence relation for $\mathfrak{B}_l^n$, we place (A2) in the recurrence relation (A1):

$$\sum_{l=0}^{n} \binom{k}{l} b_0{}^{k-l} \mathfrak{B}_l^n = \sum_{s=0}^{n} b_{n-s} \sum_{t=0}^{s} \binom{k-1}{t} b_0{}^{k-1-t} \mathfrak{B}_t^s$$
$$= \sum_{t=0}^{n} \binom{k-1}{t} b_0{}^{k-t} \mathfrak{B}_t^n + \sum_{s=0}^{n-1} \sum_{j=1}^{s+1} b_{n-s} \binom{k-1}{j-1} b_0{}^{k-j} \mathfrak{B}_{j-1}^s.$$

Since b_0 is arbitrary, we can equate its coefficients from both sides of the last expression. For $b_0{}^{k-l}$ term

$$\binom{k}{l} \mathfrak{B}_l^n = \binom{k-1}{l} \mathfrak{B}_l^n + \binom{k-1}{l-1} \sum_{s=l-1}^{n-1} b_{n-s} \mathfrak{B}_{j-1}^s.$$

Now $\binom{k}{l} - \binom{k-1}{l} = \binom{k-1}{l-1}$ so that

$$\mathfrak{B}_l^n = \sum_{s=l-1}^{n-1} b_{n-s} \mathfrak{B}_{l-1}^s \tag{A3}$$

for $l \geq 1$ and $n \geq 1$. For convenience we write some important terms

$$\mathfrak{B}_0^0 = 1, \quad \mathfrak{B}_n^n = b_1{}^n$$
$$\mathfrak{B}_0^n = 0 \quad \text{for } n \geq 1,$$
$$\mathfrak{B}_l^n = 0 \quad \text{for } l > n,$$
$$\mathfrak{B}_1^n = b_n,$$
$$\mathfrak{B}_2^n = \sum_{j=1}^{n-1} b_{n-j} b_j,$$
$$\mathfrak{B}_{n-1}^n = (n-1) b_1{}^{n-2} b_2,$$
$$\mathfrak{B}_{n-2}^n = (n-2) b_1{}^{n-3} b_3 + (n-2)(n-3) b_1{}^{n-4} b_2{}^2/2!.$$

APPENDIX B: EVALUATION OF A_k^n AND $\mathfrak{a}_j^k$

Let

$$S^n \equiv (1-\lambda)(5-\lambda)\ldots(4n-\lambda) = \sum_{k=0}^{n} A_k^n \lambda^k$$

for $n = 1, 2, 3, \cdots$. Expanding S^n for first few terms, we find

$$A^k_k = (-1)^k A^k_0 = 1\cdot 5\cdot 9\cdots(4k-3), \tag{B1a}$$

$$A^k_{k-1} = -(1 + 5 + 7 + 11\cdots), \tag{B1b}$$

$$A^k_{k-2} = (1\cdot 5 + 1\cdot 7 + 1\cdot 11 + 5\cdot 7 + 5\cdot 11 + 7\cdot 11 + \cdots), \tag{B1c}$$

$$A^k_{k-3} = -(1\cdot 5\cdot 7 + 1\cdot 5\cdot 11 + 1\cdot 7\cdot 11 + 5\cdot 7\cdot 11 + \cdots), \tag{B1d}$$

$$A^k_{k-4} = (1\cdot 5\cdot 7\cdot 11 + \cdots). \tag{B1e}$$

It is interesting that A^j_{k-j} can be thought of as the sum over all possible distinct products of the set of numbers $\{1, 5, 7, \cdots, (4k-3)\}$ when take j at a time.

Using $S^n = (4n - 3 - \lambda)S^{n-1}$ and equating coefficients of λ^k, we obtain the following recurrence relations:

$$A^n_0 = (4n-3)A^{n-1}_0 \quad \text{for } k = 0, \tag{B2a}$$

$$A^n_k = (4n-3)A^{n-1}_k - A^{n-1}_{k-1} \quad \text{for } n > k > 0, \tag{B2b}$$

$$A^n_n = -A^{n-1}_{n-1} \quad \text{for } k = n. \tag{B2c}$$

Here (B2a) and (B2c) give (B1a). Using (B2b) and summing the resulting series, we find

$$A^k_{k-1} = (-1)^{k-1}[k + 4k(k-1)/2!],$$

$$A^k_{k-2} = (-1)^{k-2}[5k(k-1)/2! + 44k(k-1)(k-2)/3! + 48k(k-1)(k-2)(k-3)/4!].$$

The form of these expressions for A^k_{k-j} indicates that we may write

$$A^k_{k-j} = (-1)^{k-j} \sum_{l=j}^{2j} \binom{k}{l} \mathfrak{A}^j_l. \tag{B3}$$

Here $\mathfrak{A}^j_l = 0$ if $l < j$ or $l > 2j$. The relation (B1a) shows

$$\mathfrak{A}^0_0 = 1, \quad \mathfrak{A}^k_k = 1\cdot 5\cdot 9\cdots(4k-3).$$

We can obtain a recurrence relation for $\mathfrak{A}^j_l$ if we put (B3) into (B2b)

$$\sum_{l=j}^{2j} \binom{k}{l} \mathfrak{A}^j_l = (4k-3) \sum_{l=j-1}^{2j-2} \binom{k-1}{l} \mathfrak{A}^{j-1}_l + \sum_{l=j}^{2j} \binom{k-1}{l} \mathfrak{A}^{j-1}_l.$$

Bringing the second term on the right over to the left-hand side of the equation and shifting the sum by 1 in the first sum on the left, we have

$$\sum_{l=j}^{2j} \binom{k-1}{l-1} \mathfrak{A}^j_l = (4k-3) \sum_{l=j}^{2j} \binom{k-1}{l-1} \mathfrak{A}^{j-1}_{l-1}. \tag{B4}$$

By induction we will show that

$$\mathfrak{A}^j_s = (4s-3)\mathfrak{A}^{j-1}_{s-1} + (4s-4)\mathfrak{A}^{j-1}_{s-2} \tag{B5}$$

for $2j \geq s \geq j$. Let $k = s$ where $2j \geq s > j$; then the sum over l in (B4) only goes up to s since $\binom{s-1}{l-1} = 0$ if $l > s$. Taking out the term $l = s$ in (B3), we have

$$\sum_{l=j}^{s-1} \binom{s-1}{l-1} \mathfrak{A}^j_l + \mathfrak{A}^j_s = (4s-3) \sum_{l=j}^{s-1} \mathfrak{A}^{j-1}_{l-1} \binom{s-1}{l-1} + (4s-3)\mathfrak{A}^{j-1}_{s-1}$$

or

$$\mathfrak{A}^j_s = (4s-3)\mathfrak{A}^{j-1}_{s-1} + \sum_{l=j}^{s-1} \binom{s-1}{l-1} [(4s-3)\mathfrak{A}^{j-1}_{l-1} - \mathfrak{A}^j_l].$$

The sum of the second term on the right only goes up to $s-1$ and we have assumed that (B5) is true for all l up to $s-1$ in order to show, by induction, that (B5) is true for $l = s$. Replacing $\mathfrak{A}^j_l$ in the last expression by (B5) and taking the term $l = s-1$ out of the resulting sum, we find

$$\begin{aligned}\mathfrak{A}^j_s &= (4s-3)\mathfrak{A}^{j-1}_{s-1} + \sum_{l=j}^{s-1} \binom{s-1}{l-1}[4(s-l)\mathfrak{A}^{j-1}_{l-1} - (4l-4)\mathfrak{A}^{j-1}_{l-2}] \\ &= (4s-3)\mathfrak{A}^{j-1}_{s-1} + 4(s-1)\mathfrak{A}^{j-1}_{s-2} \\ &\quad + 4\sum_{l=j}^{s-2} \binom{s-1}{l-1}(s-l)\mathfrak{A}^{j-1}_{l-1} - 4\sum_{l=j}^{s-1} \binom{s-1}{l-1}(l-1)\mathfrak{A}^{j-1}_{l-2}.\end{aligned}$$

By shifting the last sum by 1 it can be seen that the last two sums cancel each other. Some important values of $\mathfrak{A}^j_s$ are given in Table II. In a similar fashion we can expand

$$(3-\lambda)(7-\lambda)\cdots(4n-1-\lambda) = \sum_{k=0}^{n} \bar{A}^n_k \lambda^k.$$

Here $\bar{A}^n_k$ satisfy the following recurrence relation:

$$\bar{A}^n_k = (4n-1)\bar{A}^{n-1}_k - \bar{A}^{n-1}_{k-1}$$

for $n > k > 0$. Also

$$\bar{A}^n_0 = 3\cdot 7\cdot 11\cdots(4n-1)$$

$$\bar{A}^n_n = (-1)^n.$$

Again we can write $\bar{A}^n_k$ as

$$\bar{A}^k_{k-j} = (-1)^{k-j} \sum_{l=j}^{2j} \binom{k}{l} \bar{\mathfrak{A}}^j_l,$$

where

$$\bar{\mathfrak{A}}^j_s = (4s-1)\bar{\mathfrak{A}}^{j-1}_{s-1} + (4s-4)\bar{\mathfrak{A}}^{j-1}_{s-2}$$

for $2j > s > j$. Also $\bar{\mathfrak{A}}^j_s = 0$ if $s > j$ or $s > 2j$. See Table III for some value of $\bar{\mathfrak{A}}^j_s$.

TABLE II. Values of $\mathfrak{A}^j_{j+k}$.

j/k	0	1	2	3	4
0	1				
1	1	4			
2	5	44	48		
3	45	632	1 520	960	
4	585	11 464	44 560	60 480	26 880

TABLE III. Values of $\bar{\mathfrak{A}}^j_{j+k}$.

j/k	0	1	2	3	4
0	1				
1	3	4			
2	21	68	48		
3	231	1 272	2 000	960	
4	3 465	27 864	65 200	73 920	26 880

APPENDIX C: EVALUATION OF $Z^n_k(x)$

Starting with the definition of $Z^n_k(x)$ given by (20) and shifting the sum over j by k, we obtain

$$Z^n_k(x) = D^n \sum_{j=0}^{k} (-1)^k[(k+j)!]^{-1}\mathfrak{A}^k_{k+j} x^j D^{j+k}(\cos\sqrt{x}), \tag{C1}$$

where $D^n \equiv d^n/dx^n$. Carrying through the nth differentiation, we have

$$Z^n_k(x) = \sum_{j=0}^{k} (-1)^k[(k+j)!]^{-1} \times \mathfrak{A}^k_{k+j} \sum_{s=0}^{n} \binom{n}{s} (j-s+1)_s x^{j-s} D^{k+n+j-s}(\cos\sqrt{x}). \tag{C2}$$

Here we have made use of the factorial function $(a)_n = a(a+1)\cdots(a+n-1)$,[6] and

$$D^s(x^j) = (j-s+1)_s x^{j-s},$$

$$D^n(UV) = \sum_{s=0}^{n} \binom{n}{s} D^s(U) D^{n-s}(V).$$

Setting $t = j-s$, we can write (C2) as

$$Z_k^n = \sum_{t=0}^{k} G_t^{nk} x^t D^{k+n+t}(\cos\sqrt{x}), \tag{C3}$$

where

$$G_t^{nk} = \sum_{j=t}^{k} (-1)^k [(k+j)!]^{-1} \mathfrak{A}_{k+j}^{k} \binom{n}{j-t} (t+1)_{j-t}. \tag{C4}$$

By shifting the sum over t by $k+n$, (C3) becomes

$$Z_k^n = \sum_{l=k+n}^{2k+n} G_{l-k-n}^{nk} x^{l-k-n} D^l(\cos\sqrt{x}). \tag{C5}$$

Now

$$D^n F(\sqrt{x}) = \sum_{j=0}^{n-1} \frac{(-1)^j (n-j)_{2j} F^{(n-j)}(\sqrt{x})}{j!(2\sqrt{x})^{n+j}}$$

so that

$$D^l(\cos\sqrt{x}) = \sum_{j=0}^{l-1} \frac{(-1)^j (l-j)_{2j} \frac{1}{2}[(i)^{l-j} e^{i\sqrt{x}} + (-i)^{l-j} e^{-i\sqrt{x}}]}{j!(2\sqrt{x})^{n+j}}. \tag{C6}$$

Placing this into (C5) and letting $s = l-j$, we have

$$Z_k^n = \sum_{s=1}^{2k+n} \frac{H_s^{nk} \frac{1}{2}[(i)^s e^{i\sqrt{x}} + (-i)^s e^{-i\sqrt{x}}]}{(\sqrt{x})^{2k+2n-s}}, \tag{C7}$$

where

$$H_s^{nk} = \sum_{l=k+n}^{2k+n} (-1)^{l-s} 2^{-2l+s} [(l-s)!]^{-1} (s)_{2l-2s} G_{l-k-n}^{nk}. \tag{C8}$$

TABLE IV. Values of H_s^{0k}.

s|k	1	2	3	4
1	0	1/4	0	−19/32
2	−1/2	−1/4	0	+19/32
3		1/6	−1/8	−19/48
4		1/8	1/8	+19/96
5			−1/12	−17/480
6			−1/48	−5/48
7				+1/48
8				1/384

Note that when $s > k+n$ the sum over l in H_s^{nk} starts at $l = s$. By splitting the sum over s, in (C7), into even and odd parts, (C7) becomes

$$Z_k^n = \cos\sqrt{x} \sum_{l=1} (-1)^l H_{2l}^{nk} x^{l-k-n} + \sqrt{x} \sin\sqrt{x} \sum_{t=1} (-1)^l H_{2t-1}^{nk} x^{t-k-n}.$$

Here the upper limits of the sums over l and t are $(k+n)/2$ if n is even. When n is odd, the upper limit of l is $k + (n-1)/2$ and that of t is $k + (n+1)/2$.

The values of H_s^{nk} can be found using (C8) once the values of G_t^{nk} are known. The values of G_t^{nk} can be found using (C4) with the aid of Table II for the values of $\mathfrak{A}_{k+j}^{k}$. Table IV gives the most important values of H_s^{0k} needed to find the first few coefficients b_n.

[1] R. Vawter, Phys. Rev. **174**, 749 (1968).
[2] F. C. Auluck and D. D. Kothari, Proc. Camb. Philos. Soc. **41**, 175 (1945).
[3] P. Dean, Proc. Camb. Philos. Soc. **62**, 277 (1966).
[4] T. E. Hull and R. S. Julius, Can. J. Phys. **34**, 914 (1956).
[5] J. S. Baijal and K. K. Singh, Prog. Theor. Phys. **14**, 214 (1955).
[6] M. Abramowitz and I. A. Stegun, *Handbook Mathematical Functions* (Dover, New York, 1965).

Griffiths' inequalities for Ashkin–Teller model

Ching Tsung Lee *

Department of Physics, Alabama A&M University, Normal, Alabama 35762
(Received 26 March 1973)

The two Griffiths' inequalities for the correlation functions of Ising ferromagnets and two others added by Kelly and Sherman and by Sherman are extended to what we call generalized Ashkin–Teller model. In this model we consider a system of N particles; each can exist in r possible states. Let a collection of *pairs* of particles be represented by a graph with particles as vertices and *pairs* of particles as edges. The "many-body interaction" among a cluster of particles represented by such a graph $G(A)$ is $-J_A (J_A \geq 0)$ when the particles in each connected component of $G(A)$ all exist in the same state; it is 0 otherwise. For the special case with $r = 2$ and two-body interactions only, the Ashkin–Teller model is equivalent to Ising model. Therefore, what we present in this paper can be considered as yet another way of proving the original correlation inequalities for Ising ferromagnets with two-body interactions. We have also discovered another new inequality, namely $\langle\delta^A\rangle\langle\delta^{ABRS}\rangle + \langle\delta^{AB}\rangle\langle\delta^{ARS}\rangle - \langle\delta^{AR}\rangle\langle\delta^{ABS}\rangle - \langle\delta^{AS}\rangle\langle\delta^{ABR}\rangle \geq 0$.

1. INTRODUCTION

In 1943 Ashkin and Teller[1] considered a two-dimensional square lattice occupied by four kinds of atoms; nearest neighbors interact with only two distinct potential energies, one between like and one between unlike atoms. This was later generalized by Potts[2] and by Kihara, Midzuno and Shizume.[3] They considered r possible states for each lattice site. We will call this "Ashkin–Teller model".

Recently, Griffiths[4] obtained remarkable inequalities for the correlation functions of Ising ferromagnets with two-body interactions. These inequalities were subsequently generalized by Kelly and Sherman[5] to systems with interactions involving an arbitrary number of spins, and by Griffiths[6] to systems with arbitrary spins. Ginibre[7] further extended the second inequality.

In this paper we will prove that the two generalized Griffiths' inequalities and two others added by Kelly and Sherman[5] and by Sherman[8] are also true in the generalized Ashkin–Teller model with interactions involving arbitrary "clusters" of particles. By a cluster of particles we mean a collection of *pairs* of particles rather than a group of particles. We will present the two generalized Griffiths' inequalities as Theorem I and Theorem II; the one added by Kelly and Sherman as Theorem III; and that by Sherman as Theorem IV. We have also discovered a new inequality which we present as a corollary to Theorem II.

2. GENERALIZED ASHKIN-TELLER MODEL

Consider a system of N particles identified by the index $i = 1, 2, \ldots, N$. The location of the ith particle will also be denoted by i. There will not be a lattice, but only a set of sites without any restriction on their geometrical arrangement or on the dimension of the space. Each particle can exist in r possible states denoted by $\rho_i = 1, 2, \ldots, r$. A pair of particles will be denoted by (i,j) with $i \neq j$. There are $N(N-1)/2$ such pairs. Let M be the whole set of these pairs. We will consider the i's as vertices and the (i,j)'s as edges. Then corresponding to each subset A of M, there is a graph denoted by $G(A)$. (Note that an isolated point will not be considered as part of a graph.) For each such graph $G(A)$, we will define a number $d(A)$ as follows:

$$d(A) = v(A) - c(A), \tag{2.1}$$

where $v(A)$ is the number of vertices of $G(A)$ and $c(A)$ is the number of connected components of $G(A)$.

For each (i,j), we define a Kronecker delta

$$\delta(\rho_i, \rho_j) = \begin{cases} 1 & \text{for } \rho_i = \rho_j \\ 0 & \text{for } \rho_i \neq \rho_j . \end{cases} \tag{2.2}$$

For each subset A of M, we also define

$$\delta^A = \prod_{(i,j)\in A} \delta(\rho_i, \rho_j). \tag{2.3}$$

Since $[\delta(\rho_i, \rho_j)]^2 = \delta(\rho_i, \rho_j)$, we have

$$\delta^A \delta^B = \delta^{A\cup B}. \tag{2.4}$$

For convenience, "$A\cup B$" will be abbreviated as "AB", "$A\cup B\cup C$" as "ABC", etc. We note in passing that $\delta^A\delta^A = \delta^A$.

Let γ denote a configuration, i.e., a possible assignment of the values of all the ρ's; and let $(\delta^A)_\gamma$ be the value of δ^A in the configuration γ. We see that $(\delta^A)_\gamma = 1$ only when the ρ's in the same connected component of $G(A)$ have the same value in the configuration γ; it is 0 otherwise.

The generalized Ashkin–Teller model can be described by the Hamiltonian of the system whose value at the configuration γ is

$$H_\gamma = -\sum_{A\subset M} J_A (\delta^A)_\gamma, \tag{2.5}$$

where $J_A \geq 0$, $J_\phi = 0$, and the summation is over all the subsets of M.

The partition function of the system can be written as

$$Z = \sum_\gamma \exp(-\beta H_\gamma), \tag{2.6}$$

where $\beta = (\kappa T)^{-1}$, κ being Boltzmann's constant and T the (absolute) temperature, and the summation is over all the r^N possible configurations. The average value of δ^A can be calculated by the following formula:

$$\langle\delta^A\rangle = Z^{-1}\sum_\gamma (\delta^A)_\gamma \exp(-\beta H_\gamma). \tag{2.7}$$

3. MAIN THEOREMS

The generalized Griffiths' inequalities for our generalized Ashkin–Teller model can be written as the following two theorems:

Theorem I:

$$1 \geq \langle\delta^R\rangle \geq r^{-d(R)}, \quad \text{for all } R \subset M. \tag{3.1}$$

Theorem II:

$$\langle\delta^{RS}\rangle - \langle\delta^{R}\rangle\langle\delta^{S}\rangle \geq 0, \quad \text{for all } R, S \subset M. \tag{3.2}$$

On account of the identity

$$\frac{1}{\beta}\frac{\partial\langle\delta^{R}\rangle}{\partial J_S} = \langle\delta^{RS}\rangle - \langle\delta^{R}\rangle\langle\delta^{S}\rangle, \tag{3.3}$$

we see that Theorem II has the following physical meaning: The correlation of the states of any cluster of particles $\langle\delta^R\rangle$ can never decrease when any of the interaction constants J_S is increased. Noting that when all the J's vanish we have the equality $\langle\delta^R\rangle = r^{-d(R)}$, we see that Theorem I is a natural consequence of Theorem II.

Theorem II is actually a special case of a more general inequality which we will call

Theorem II':

$$\langle\delta^{A}\rangle\langle\delta^{ARS}\rangle - \langle\delta^{AR}\rangle\langle\delta^{AS}\rangle \geq 0, \quad \text{for all } A, R, S \subset M. \tag{3.4}$$

In Sec. 6 we will prove Theorem II'. When we let $A = \phi$, we will obtain Theorem II automatically. However, the physical meaning of Theorem II' is unknown.

As a byproduct of proving Theorem II', we obtain a new inequality:

$$\langle\delta^{A}\rangle\langle\delta^{ABRS}\rangle + \langle\delta^{AB}\rangle\langle\delta^{ARS}\rangle - \langle\delta^{AR}\rangle\langle\delta^{ABS}\rangle - \langle\delta^{AS}\rangle\langle\delta^{ABR}\rangle \geq 0, \quad \text{for all } A, B, R, S \subset M. \tag{3.5}$$

This inequality will be identified as a corollary to Theorem II'.

As observed earlier, Theorem II implies that none of the moments $\langle\delta^R\rangle$ can decrease when J_S is increased. One has a strong feeling that $\langle\delta^S\rangle$ should increase at a faster rate than any of the other $\langle\delta^R\rangle$. That this is so is confirmed by

Theorem III:

$$\frac{1}{\beta}\frac{\partial\langle\delta^{S}\rangle}{\partial J_S} - \frac{1}{\beta}\frac{\partial\langle\delta^{R}\rangle}{\partial J_S} = \langle\delta^{S}\rangle - \langle\delta^{S}\rangle^2 - \langle\delta^{RS}\rangle + \langle\delta^{R}\rangle\langle\delta^{S}\rangle \geq 0, \quad \text{for all } R, S \subset M. \tag{3.6}$$

The upper bound for the rate of change of $\langle\delta^R\rangle$ with respect to J_S is given by

Theorem IV:

$$\frac{\beta}{2}[1 + \langle\delta^{RS}\rangle^2 - \langle\delta^{R}\rangle^2 - \langle\delta^{S}\rangle^2] \geq \frac{\partial\langle\delta^{R}\rangle}{\partial J_S}, \quad \text{for all } R, S \subset M. \tag{3.7}$$

4. PROOF OF THEOREM I

Lemma I:

$$d(A) + d(B) \geq d(AB). \tag{4.1}$$

Proof: When we put two graphs $G(A)$ and $G(B)$ together to obtain $G(AB)$, it is possible that the number of connected components is reduced; this will happen when components of $G(A)$ and components of $G(B)$ overlap. So we have

$$c(AB) \leq c(A) + c(B). \tag{4.2}$$

It is also possible that the number of vertices is reduced; this will happen when vertices of $G(A)$ and vertice of $G(B)$ overlap. So we also have

$$v(AB) \leq v(A) + v(B). \tag{4.3}$$

But we can see that corresponding to each reduction in c, there must be at least one common vertex in $G(A)$ and $G(B)$; there may be more. Hence the reduction in v is always greater than or equal to the reduction in c, i.e.,

$$v(A) + v(B) - v(AB) \geq c(A) + c(B) - c(AB), \tag{4.4}$$

or

$$[v(A) - c(A)] + [v(B) - c(B)] \geq [v(AB) - c(AB)]. \tag{4.5}$$

And by definition Lemma I follows. End of proof of Lemma I.

Theorem I:

$$1 \geq \langle\delta^{R}\rangle \geq r^{-d(R)}. \tag{3.1}$$

Proof: The value of δ^R is either 1 or 0; so the first part of the above inequality is obviously true. To prove the second part, we see that for any $R \subset M$ we have

$$\begin{aligned} Z\langle\delta^{R}\rangle &= \sum_\gamma (\delta^R)_\gamma \exp\left(\beta \sum_{A\subset M} J_A(\delta^A)_\gamma\right) \\ &= \sum_\gamma (\delta^R)_\gamma \sum_{k=0}^{\infty} \frac{\beta^k}{k!}\left(\sum_{A\subset M} J_A(\delta^A)_\gamma\right)^k \\ &= \sum_{k=0}^{\infty} \frac{\beta^k}{k!} \sum_{A_1\subset M} J_{A_1} \cdots \sum_{A_k\subset M} J_{A_k} \sum_\gamma (\delta^{A_1\cdots A_k R})_\gamma. \end{aligned} \tag{4.6}$$

But we have

$$\sum_\gamma (\delta^A)_\gamma = r^{N-d(A)}; \tag{4.7}$$

and by Lemma I,

$$d(A_1\cdots A_k R) \leq d(A_1\cdots A_k) + d(R). \tag{4.8}$$

Therefore, we have

$$\begin{aligned} \sum_\gamma (\delta^{A_1\cdots A_k R})_\gamma &= r^{N-d(A_1\cdots A_k R)} \\ &\geq r^{N-d(A_1\cdots A_k)-d(R)} = r^{-d(R)} \sum_\gamma (\delta^{A_1\cdots A_k})_\gamma. \end{aligned} \tag{4.9}$$

Using (4.9) in (4.6), we obtain

$$Z\langle\delta^{R}\rangle \geq r^{-d(R)} Z; \tag{4.10}$$

and the second part of Theorem I follows. End of proof of Theorem I.

5. PROOF OF THE COROLLARY TO THEOREM II'

Because of the way we will prove Theorem II', it is more convenient to present the proof of the corollary before the proof of Theorem II'. We will just assume the truth of Theorem II' here.

Lemma II: Let $a, b, c,$ and d be all positive numbers and $a \geq c$, $a \geq d$, $ab \geq cd$; then $a + b \geq c + d$.

Proof: Without losing generality, let us assume that $c \geq d$. We consider two cases.

Case I: $b \geq d$.

Combining $a \geq c$ and $b \geq d$, we get $a + b \geq c + d$.

Case II: $b < d$.

$ab \geq cd \Rightarrow ab - bd \geq cd - bd \Rightarrow (a - d)b \geq (c - b)d$. But $b < d$; so we have $a - d > c - b \Rightarrow a + b > c + d$. End of proof of Lemma II.

Corollary:

$$\langle\delta^A\rangle\langle\delta^{ABRS}\rangle + \langle\delta^{AB}\rangle\langle\delta^{ARS}\rangle - \langle\delta^{AR}\rangle\langle\delta^{ABS}\rangle - \langle\delta^{AS}\rangle\langle\delta^{ABR}\rangle \geq 0. \tag{3.5}$$

Proof: Let $a = \langle\delta^A\rangle\langle\delta^{ABRS}\rangle$, $b = \langle\delta^{AB}\rangle\langle\delta^{ARS}\rangle$, $c = \langle\delta^{AR}\rangle\langle\delta^{ABS}\rangle$, and $d = \langle\delta^{AS}\rangle\langle\delta^{ABR}\rangle$. From the first part of Theorem I, a, b, c, and d are all positive numbers. Assuming the truth of Theorem II', we have $a \geq c$ and $a \geq d$. In addition, since

$$ab = \langle\delta^A\rangle\langle\delta^{ABRS}\rangle\langle\delta^{AB}\rangle\langle\delta^{ARS}\rangle = \{\langle\delta^A\rangle\langle\delta^{ARS}\rangle\}\{\langle\delta^{AB}\rangle\langle\delta^{ABRS}\rangle\}, \tag{5.1}$$

$$cd = \langle\delta^{AR}\rangle\langle\delta^{ABS}\rangle\langle\delta^{AS}\rangle\langle\delta^{ABR}\rangle = \{\langle\delta^{AR}\rangle\langle\delta^{AS}\rangle\}\{\langle\delta^{ABR}\rangle\langle\delta^{ABS}\rangle\}, \tag{5.2}$$

we see that the two factors of ab is each greater than or equal to the two factors of cd, respectively. So we also have $ab \geq cd$. All the conditions of Lemma II are satisfied. By this Lemma, we have $a + b \geq c + d$. This is just the content of the Corollary. End of proof of the Corollary.

6. PROOF OF THEOREM II'

Definition I:

$$g(A, B) = d(A) + d(B) - d(AB). \tag{6.1}$$

This is the reduction in d when we put $G(A)$ and $G(B)$ together. By Lemma I we have $g(A, B) \geq 0$.

Lemma III: If $G(C)$ is a subgraph of $G(B)$, then

$$g(A, B) \geq g(A, C). \tag{6.2}$$

Proof: If $G(C)$ is a subgraph of $G(B)$, the intersection between $G(A)$ and $G(B)$ will be more than or equal to that already exists between $G(A)$ and $G(C)$. By the same argument presented in the proof of Lemma I, more intersection can only lead to further reduction in d; or at least remains the same. End of proof of Lemma III.

Lemma IV:

$$d(A) + d(ARS) \leq d(AR) + d(AS). \tag{6.3}$$

Proof: By Definition I, we have

$$d(A) + d(ARS) = d(A) + d(R) + d(AS) - g(R, AS), \tag{6.4}$$

$$d(AR) + d(AS) = d(A) + d(R) - g(R, A) + d(AS). \tag{6.5}$$

And by Lemma III, we have

$$g(R, AS) \geq g(R, A) \tag{6.6}$$

since $G(A)$ is a subgraph of $G(AS)$. Using (6.6) we see that the right-hand side of (6.4) is less than or equal to that of (6.5); then the left hand sides of these two equations will give Lemma IV. End of proof of Lemma IV.

We note in passing that Lemma I is actually a special case of Lemma IV when $G(A)$ is a null graph.

Definition II:

$$X_A = \exp(\beta J_A) - 1. \tag{6.7}$$

By this definition we have $X_A = 0$ when $J_A = 0$; and X_A increases monotonically as J_A increases. We also have the following expression:

$$\exp[\beta J_A(\delta^A)_\gamma] = 1 + (\delta^A)_\gamma X_A. \tag{6.8}$$

Theorem II':

$$\langle\delta^A\rangle\langle\delta^{ARS}\rangle - \langle\delta^{AR}\rangle\langle\delta^{AS}\rangle \geq 0. \tag{3.4}$$

Proof: We will prove this theorem by mathematical induction in the following steps:

1. Let all the J's be zero: In this case we have

$$Z = r^N, \tag{6.9}$$

the total number of configurations. Therefore,

$$\langle\delta^A\rangle = Z^{-1}\sum_\gamma(\delta^A)_\gamma = Z^{-1}r^{N-d(A)} = r^{-d(A)}. \tag{6.10}$$

Hence we have

$$\langle\delta^A\rangle\langle\delta^{ARS}\rangle - \langle\delta^{AR}\rangle\langle\delta^{AS}\rangle = r^{-d(A)}r^{-d(ARS)} - r^{-d(AR)}r^{-d(AS)} = r^{-[d(A)+d(ARS)]} - r^{-[d(AR)+d(AS)]}. \tag{6.11}$$

Using Lemma IV in (6.11), we will obtain (3.4). Hence Theorem II' is true when all the J's are zero. The truth of the Corollary for this particular case follows.

2. Suppose Theorem II' and the Corollary are true when a certain number of the J's are not zero (including the case when all the J's are zero), we will prove that Theorem II' is again true when an additional J_B is not zero.

Let the partition function before the addition of nonvanishing J_B be written as

$$Z = \sum_\gamma Z_\gamma; \tag{6.12}$$

and that after the addition be written as

$$Z' = \sum_\gamma Z_\gamma \exp[\beta J_B(\delta^B)_\gamma] = \sum_\gamma Z_\gamma[1 + (\delta^B)_\gamma X_B], \tag{6.13}$$

where we have used (6.8). Let the average value of δ^A be denoted by $\langle\delta^A\rangle$ when the partition function is Z; and by $\langle\delta^A\rangle'$ when the partition function is Z'. Then we have

$$\begin{aligned} Z'\langle\delta^A\rangle' &= \sum_\gamma(\delta^A)_\gamma Z_\gamma[1 + (\delta^B)_\gamma X_B] \\ &= \sum_\gamma(\delta^A)_\gamma Z_\gamma + X_B\sum_\gamma(\delta^{AB})_\gamma Z_\gamma \\ &= Z\langle\delta^A\rangle + Z\langle\delta^{AB}\rangle X_B. \end{aligned} \tag{6.14}$$

Using (6.14) repeatedly, we obtain

$$\begin{aligned} Z'^2\{&\langle\delta^A\rangle'\langle\delta^{ARS}\rangle' - \langle\delta^{AR}\rangle'\langle\delta^{AS}\rangle'\} \\ &= Z^2\{[\langle\delta^A\rangle + \langle\delta^{AB}\rangle X_B][\langle\delta^{ARS}\rangle + \langle\delta^{ABRS}\rangle X_B] \\ &\quad - [\langle\delta^{AR}\rangle + \langle\delta^{ABR}\rangle X_B][\langle\delta^{AS}\rangle + \langle\delta^{ABS}\rangle X_B]\} \\ &= Z^2\{[\langle\delta^A\rangle\langle\delta^{ARS}\rangle - \langle\delta^{AR}\rangle\langle\delta^{AS}\rangle] \\ &\quad + [\langle\delta^A\rangle\langle\delta^{ABRS}\rangle + \langle\delta^{AB}\rangle\langle\delta^{ARS}\rangle \\ &\quad - \langle\delta^{AR}\rangle\langle\delta^{ABS}\rangle - \langle\delta^{AS}\rangle\langle\delta^{ABR}\rangle]X_B \\ &\quad + [\langle\delta^{AB}\rangle\langle\delta^{ABRS}\rangle - \langle\delta^{ABR}\rangle\langle\delta^{ABS}\rangle]X_B^2\}. \end{aligned} \tag{6.15}$$

Assuming the truth of Theorem II′ before the addition of the nonvanishing J_B, we see that the constant term and the coefficient of the quadratic term are both nonnegative. And the truth of the Corollary implies that the coefficient of the linear term is also nonnegative. Hence we have

$$\langle\delta^A\rangle'\langle\delta^{ARS}\rangle' - \langle\delta^{AR}\rangle'\langle\delta^{AS}\rangle' \geq 0. \tag{6.16}$$

Therefore, Theorem II′ is again true when the nonvanishing J_B is added. The truth of the Corollary for this case follows again.

3. Combining step 1 and step 2, and using the argument of mathematical induction, we can complete our proof of Theorem II′.

7. PROOF OF THEOREM III

Lemma V:

$$d(RS) \geq d(R). \tag{7.1}$$

Proof: When we put graph $G(S)$ onto graph $G(R)$, the number of vertices can not decrease, i.e., $v(RS) \geq v(R)$. This is favorable to the inequality. On the other hand, for any increase in c by 1, there must be a corresponding increase in v by at least 2, i.e., the increase in c is always over compensated by the increase in v. Hence the truth of our Lemma. End of proof of Lemma V.

Lemma VI:

$$\langle\delta^R\rangle \geq \langle\delta^{RS}\rangle. \tag{7.2}$$

Proof: From (4.7) and (4.8) we have

$$Z\langle\delta^R\rangle = \sum_{k=0}^{\infty}\frac{\beta^k}{k!}\sum_{A_1\subset M}J_{A_1}\cdots\sum_{A_k\subset M}J_{A_k}\gamma^{N-d(A_1\cdots A_kR)}, \tag{7.3}$$

$$Z\langle\delta^{RS}\rangle = \sum_{k=0}^{\infty}\frac{\beta^k}{k!}\sum_{A_1\subset M}J_{A_1}\cdots\sum_{A_k\subset M}J_{A_k}\gamma^{N-d(A_1\cdots A_kRS)}. \tag{7.4}$$

And by Lemma V, we have $d(A_1\cdots A_kRS) \geq d(A_1\cdots A_kR)$. Therefore, each term of (7.3) is greater than or equal to the corresponding term of (7.4). So Lemma VI follows. End of proof of Lemma VI.

Theorem III:

$$\langle\delta^S\rangle - \langle\delta^S\rangle^2 - \langle\delta^{RS}\rangle + \langle\delta^R\rangle\langle\delta^S\rangle \geq 0. \tag{3.6}$$

Proof: Let $a = \langle\delta^S\rangle$, $b = \langle\delta^S\rangle\langle\delta^R\rangle$, $c = \langle\delta^S\rangle^2$, and $d = \langle\delta^{RS}\rangle$. We see that a, b, c, and d are all positive numbers; $a \geq c$ since $\langle\delta^S\rangle \leq 1$; $a \geq d$ by Lemma VI; and $ab = \langle\delta^S\rangle^2\langle\delta^R\rangle$, $cd = \langle\delta^S\rangle^2\langle\delta^{RS}\rangle$, again by Lemma VI we have $ab \geq cd$. Thus all the conditions for Lemma II are satisfied. Hence Theorem III follows from Lemma II. End of proof of Theorem III.

8. PROOF OF THEOREM IV

Theorem IV:

$$1 + \langle\delta^{RS}\rangle^2 - \langle\delta^R\rangle^2 - \langle\delta^S\rangle^2 \geq \frac{2}{\beta}\frac{\partial\langle\delta^R\rangle}{\partial J_S}. \tag{3.7}$$

Proof: By Theorem I we have $1 \geq \langle\delta^R\rangle$; by Lemma VI we have $\langle\delta^S\rangle \geq \langle\delta^{RS}\rangle$. Adding these two inequalities, we obtain

$$1 - \langle\delta^{RS}\rangle \geq \langle\delta^R\rangle - \langle\delta^S\rangle. \tag{8.1}$$

Similarly, we can also have

$$1 - \langle\delta^{RS}\rangle \geq \langle\delta^S\rangle - \langle\delta^R\rangle. \tag{8.2}$$

Combining (8.1) and (8.2), we obtain

$$[1 - \langle\delta^{RS}\rangle]^2 \geq [\langle\delta^R\rangle - \langle\delta^S\rangle]^2, \tag{8.3}$$

or

$$1 + \langle\delta^{RS}\rangle^2 - \langle\delta^R\rangle^2 - \langle\delta^S\rangle^2 \geq 2[\langle\delta^{RS}\rangle - \langle\delta^R\rangle\langle\delta^S\rangle]. \tag{8.4}$$

Using (3.3) in (8.4), Theorem IV follows. End of proof of Theorem IV.

9. DISCUSSION

In this paper we have proved several correlation inequalities for the generalized Ashkin–Teller model with many-body interactions. Our many-body interactions are somewhat different from ordinary ones. Usually people consider many-body interactions as corresponding to groups of particles; while we consider them as corresponding to collections of pairs of particles, or corresponding to graphs with particles as vertices and pairs of particles as edges. If we restrict ourselves to considering interactions corresponding to complete graphs only, then our many-body interactions will be the same as ordinary ones. In this sense, we may regard our type of many-body interactions as "generalized many-body interactions."

When the number of possible states of each particle is two and all the many-body interactions except the two-body ones are vanishing, then Ashkin–Teller model is equivalent to Ising model. This can be seen through the following transformation:

$$\sigma_k\sigma_l = 2\delta(\rho_k, \rho_l) - 1, \tag{9.1}$$

where the σ's are the spins in Ising model which take the values "+ 1" or "− 1". The two original Griffiths' inequalities for Ising model, i.e.,

$$\text{I.}\quad \langle\sigma_k\sigma_l\rangle \geq 0, \tag{9.2}$$

$$\text{II.}\quad \langle\sigma_k\sigma_l\sigma_m\sigma_n\rangle - \langle\sigma_k\sigma_l\rangle\langle\sigma_m\sigma_n\rangle \geq 0, \tag{9.3}$$

follow immediately from our Theorem I and Theorem II. This will be shown in the following:

Proof of I: When the graph $G(R)$ consists of a single edge only, we have

$$d(R) = v(R) - c(R) = 2 - 1 = 1. \tag{9.4}$$

Then, from Theorem I, we have

$$\langle\delta(\rho_k, \rho_l)\rangle \geq 1/2. \tag{9.5}$$

Hence

$$\langle\sigma_k\sigma_l\rangle = 2\langle\delta(\rho_k, \rho_l)\rangle - 1 \geq 0. \tag{9.6}$$

End of proof of I.

Proof of II: Let $G(R)$ and $G(S)$ be two graphs each consists of a single edge, (k, l) and (m, n), respectively. Then, using Theorem II, we have

$$\langle\delta(\rho_k, \rho_l)\delta(\rho_m, \rho_n)\rangle - \langle\delta(\rho_k, \rho_l)\rangle\langle\delta(\rho_m, \rho_n)\rangle \geq 0. \tag{9.7}$$

On the other hand, we also have

$$\langle \sigma_k \sigma_l \sigma_m \sigma_n \rangle = \langle [2\delta(\rho_k, \rho_l) - 1][2\delta(\rho_m, \rho_n) - 1] \rangle$$
$$= 4\langle \delta(\rho_k, \rho_l)\delta(\rho_m, \rho_n) \rangle - 2\langle \delta(\rho_k, \rho_l) \rangle$$
$$- 2\langle \delta(\rho_m, \rho_n) \rangle + 1, \tag{9.8}$$

$$\langle \sigma_k \sigma_l \rangle \langle \sigma_m \sigma_n \rangle = [2\langle \delta(\rho_k, \rho_l) \rangle - 1][2\langle \delta(\rho_m, \rho_n) \rangle - 1]$$
$$= 4\langle \delta(\rho_k, \rho_l) \rangle \langle \delta(\rho_m, \rho_n) \rangle - 2\langle \delta(\rho_k, \rho_l) \rangle$$
$$- 2\langle \delta(\rho_m, \rho_n) \rangle + 1. \tag{9.9}$$

Comparing the right-hand sides of (9.8) and (9.9), and using (9.7), we obtain II. End of proof of II.

From the above discussion, we can say that what we have presented in this paper can also be considered as a new way of proving the original Griffiths' inequalities for Ising model. Our methods have been mostly graph-theoretical. In our opinion, these are more elementary than what have been used so far.

ACKNOWLEDGMENT

I am grateful to Dr. C. C. Rousseau for stimulating discussions and valuable suggestions.

* Work supported in part by NASA.

[1] J. Ashkin and E. Teller, Phys. Rev. **64**, 178 (1943).
[2] R. B. Potts, Proc. Camb. Philos. Soc. **48**, 106 (1952).
[3] T. Kihara, Y. Midzuno, and T. Shizume, J. Phys. Soc. Jap. **9**, 681 (1954).
[4] R. B. Griffiths, J. Math. Phys. **8**, 478 (1967), J. Math. Phys. **8**, 484 (1967).
[5] D. G. Kelly and S. Sherman, J. Math. Phys. **9**, 466 (1968).
[6] R. B. Griffiths, J. Math. Phys. **10**, 1559 (1969).
[7] J. Ginibre, Phys. Rev. Lett. **23**, 828 (1969); Commun. Math. Phys. **16**, 312 (1970).
[8] S. Sherman, Commun. Math. Phys. **14**, 1 (1969).

Irreversible behavior of a thermally insulated system

Julio F. Fernández

Instituto Venezolano de Investigaciones Cientfficas, Caracas 101, Venezuela
(Received 7 March 1973)

The time evolution of a system coupled to a reservoir, in such a way that no energy exchange at all takes place, is examined, considering the system plus reservoir as dynamically closed. An H-theorem is proved, more precisely, it is shown quite generally that if the system and reservoir are initially uncorrelated, then the Gibbs–Jaynes entropy at time t, $S(t)$, obeys the relation $S(t) \geq S(0)$. We have previoulsy exhibited a simple model where $S(t) > S(0)$. Some relaxation times of $S(t)$, for an x–y model (system) interacting (no energy exchange) with an Ising model (reservoir), are obtained approximately. The values obtained seem reasonable. It is also shown, for this particular model, that if the motion of the system, but not the reservoir, is inverted at time T, then its Gibbs–Jaynes entropy τ, later $S^{MI}(\tau; T)$, is given by $S^{MI}(\tau; T) \simeq S(T + \tau)$, i.e., the entropy keeps evolving in time as if the motion inversion had not taken place.

I. INTRODUCTION

The solution to the approach to equilibrium problem consists in describing the time evolvement of a system of particles initially not in thermodynamic equilibrium. Physical quantities which exist in nonstatistical physics have a dynamical operator associated with it (entropy does not belong to this class of quantities). The expectation value of an operator A, as a function of time, is given by[1]

$$\langle A(t)\rangle = \mathrm{Tr}[A\rho(t)]. \tag{1}$$

For a dynamically isolated system with a Hamiltonian independent of time

$$\langle E_j|\rho(t)|E_i\rangle = \exp[-i(E_j - E_i)th^{-1}]\langle E_j|\rho(o)|E_i\rangle, \tag{2}$$

where

$$H|E_i\rangle = E_i|E_i\rangle. \tag{3}$$

Then,

$$\langle A(t)\rangle = \sum_{ij}\langle E_i|A|E_j\rangle\langle E_j|\rho(o)|E_i\rangle e^{\frac{-i(E_j - E_i)t}{\hbar}}.$$

If the set of energy levels becomes continuous, the sum in the above expression becomes an integral. The possibility for $\langle A(t)\rangle$ to approach a constant value as t increases, clearly, exists. There may exist, therefore, some characteristic time τ_A for the approach to equilibrium of $\langle A(t)\rangle$.

The following question, discussed by Blatt,[2] however, remains. Suppose the dynamically closed system suffers an inversion of the motion (for point particles, it would amount to reserving all the particle velocities instantaneously) after some time much larger than τ_A. According to the laws of dynamics,[3] both classically and quantum mechanically, the system would evolve in time as in a movie (taken when the system was running forward) run backwards. Such a time reversal of a spin system has been performed experimentally by Hahn[4] and by Rhim[5] *et al.*, with results in agreement with the predictions of dynamics. Suppose the system is thermally insulated after an extremely long time and one performs the motion inversion operation. One would not expect the system to go back to the initial nonequilibrium state, on intuitive grounds. However, if one thinks of the system as dynamically isolated one reaches the opposite conclusion. This apparent contradiction led Blatt[2] to point out that one should not think of a thermally insulated system as a dynamically isolated one. He suggested that there are interactions between the system and the surroundings which will bring the system into thermal equilibrium, i.e., $\rho(t)$ would become diagonal in the energy representation long before there is any energy flow into or out of the system. That is, according to Blatt $\rho(t)$ would become diagonal with a characteristic time τ_s, long before τ_E (the characteristic time for any energy exchange to occur). And often $\tau_A \ll \tau_s$.

The density matrix associated with thermodynamic equilibrium is given by

$$\rho = f(H), \tag{4}$$

where $f(H)$ has different forms in the microcanonical and canonical ensembles. Equation (2) shows that a dynamically isolated system, initially in a nonequilibrium condition, never satisfies Eq. (4).

The Gibbs–Jaynes entropy[6] is given by

$$S(t) = -k_B T_r[\rho_s(t) ln \rho_s(t)], \tag{5}$$

where k_B is Boltzmann's constant and $\rho_s(t)$ is the density matrix of the system at time t. Equations (4) and (5) imply

$$S(t) = S(o), \tag{6}$$

a seemingly disconcerting result, in view of the second law of thermodynamics. One may look at Eqs. (5) and (6) (or rather, the classical analogs) as a manifestation of the fact that the time evolution of the distribution function (analog of ρ) in classical physics is governed by a measure preserving transformation if the system is dynamically isolated. Simply, the volumes of the phase space cloud is a constant of the motion.

Two types of approach have been used to avoid the annoying result expressed by Eq. (6), which shall now be briefly discussed:

(a) The effect due to interactions between the thermodynamically insulated system of interest and outside systems is accounted for phenomenologically.

Clearly, if a thermally insulated system is acted on, even if ever so subtly so as not to transfer any heat, by some external system, then Eq. (6) need not be fulfilled. The transformation governing the time evolution of the distribution function for the system of interest need not be measure preserving any longer. One may mimic this effect by coarse graining repeatedly in the course of time.[7] This procedure yields the well known master equation, which breaks time reversal invariance, and produces $S(t) \geq S(o)$.

On the other hand, one may derive the master equation rigorously[8] for a closed system, thus obtaining the time evolution of the diagonal part of the density matrix if one assumes it to be *initially* diagonal. In this case, in contradistinction to the coarse graining scheme, the off-diagonal part of the density matrix plays a crucial role in a motion inversion experiment. The apparent break of time reversal invariance is not real in this case. This is so because whereas forward motion proceeds under the assumption that the density matrix is initially diagonal (allowing one to use the master equation), no such assumption can be made for the initial condition of the system after its motion is inverted.

The introduction of a transformation which does not preserve measure may also be accomplished using stochastic fields,[9] again, to try to cancel the mistake, so to speak, of considering the system dynamically isolated. To have a stochastic field $X(t)$ means to have different probabilities $X(t)$ to take all sorts of values in the course time. Now given an $X(t)$, a point in phase space will evolve along a certain trajectory. Therefore, to have a stochastic field $X(t)$ implies that a phase space point will evolve along several trajectories with a certain probability for each one of them. Clearly, one no longer has a measure preserving transformation in phase space. It is then possible to show[9] that $S(t) \geq S(o)$ for a class of stochastic fields. Such a procedure may break time reversal invariance.

(b) Thermal insulation is equated to dynamical isolation. The dynamical equations of motion are not tampered with, but Eq. (5) is modified.

Some examples of this type of treatment are ergodic theory,[10] C systems,[11] and mixing flow.[10,12] We shall just say a word about mixing flow, but the basic comments apply to the ergodic theory and C system approaches just as well.

Crudely speaking, a system is said to follow mixing flow if any set of points of finite volume on an energy surface in phase space evolves in time in such a way that in the $t \to \infty$ limit it covers the whole energy surface uniformly in a coarse grained way. The coarse graining is completely arbritary, as long as the energy surface is divided into cells of finite volume. It is intuitively clear that mixing flow implies ergodicity and that the converse is not true. Clearly, if one looks at the distribution function in a coarse grained way after $t \to \infty$, then $S(t = \infty) \geq S(o)$ if the system obeys mixing flow. One such system is a set of hard spheres in a box, as has been shown by Sinai.[13] Time reversal invariance is not broken, however.

It may be worth remarking that in all of these treatments one deals with dynamically isolated system, strictly adhering to the laws of dynamics; therefore, one cannot escape the existence of Poincaré recurrence times if one treats *finite quantum* mechanical systems. Indeed, Eq. (2) shows that $\rho(t)$ will recur in time if the number of states initially occupied is finite, which is the case for finite systems with bound energies. This is to be contrasted with *finite classical* systems, where as Lebowitz[12] has pointed out there is *no* recurrence time if there is mixing flow and if the system is described by a set of finite volume in phase space.

We shall shortly point out some connections between the approach we intend to follow and the methods of attack described under parts "a" and "b".

A thermally insulated system will betreated here as *interacting* with another system (reservoir), but taking into account the special nature of the reservoir (thermally insulating) such that no energy at all goes in or out of it. To meet this condition we shall take

$$[H_R, H] = 0, \tag{7}$$

where H_R is the reservoir Hamiltonian and H is the Hamiltonian of the whole assembly (system plus reservoir).

To be more specific, let

$$H = H_S + H_I + H_R,$$

where H_S, H_R, and H_I are the Hamiltonian of the system, the reservoir, and interaction between system and reservoir, respectively.

Notice that Eq. (7) implies *not* just that the reservoir gets zero *mean* energy. It does imply that there is no energy flow into or out of the reservoir at all regardless of the state of the system and the reservoir. This is a radical idealization of thermal insulation, but it is more realistic than equating it to dynamical isolation.

Systems interacting with reservoirs have been the subject of many papers.[14] However, energy exchange was readily allowed.

We have previously shown,[15] for a very simple model satisfying Eq. (7), that the entropy of the system [defined by Eq. (5)] satisfies $S(t = \infty) > S(0)$ if the system density matrix is initially nondiagonal in the representation of H_s, and if the system and reservoir *(infinite)* are initially uncorrelated. It was also shown that there is no recurrence time for such a system (finite), and that if the system but *not* the reservoir is inverted at time T, then the entropy of the system at time t afterwards is given by $S(T + t)$, i.e., the entropy of the system, *for the model of Ref. (15)* keeps evolving as if no motion inversion had taken place. Thus, time reversal invariance is broken. No statistical assumptions were made. The time evolution of the density matrix for the system ρ_s was obtained from the dynamics of the dynamically closed assembly (system + reservoir).

An examination of the time evolution of ρ_s yields the following picture (see the general equation (14)) the reservoir acts as a source of stochastic *static* fields acting on the system, and, of course, the time evolution of the system is governed by a transformation which does not preserve measure. That the fields turn out to be static is a manifestation of thermal insulation.

The relation between this treatment and those described under part "a" becomes immediately obvious. This approach may be thought of, therefore, as an attempt at first principles justification of such phenomenological treatments.

Now, in mixing flow one may crudely think of a phase space could be come so filamentary in the $t \to \infty$ limit that it fills energy surfaces uniformly in a coarse grained way. The effect of the external thermally insulating reservoir is, presumably, to smear the cloud filaments so that there is no difference between the fine grained and coarse grained phase space distribution functions, and consequently the fine grained entropy, Eq. (5), and the coarse grained version should agree.

Our treatment supports the mixing flow approach in the following academic point. The whole idea of mixing flow rests on the assumption that the system is not a pure state, i.e., the state of a classical system is represented in phase space by a set of points of nonzero

volume. Now in Ref. (15) it was shown that, at least for a particular initial pure state of the reservoir, and any pure state of the system (except an eigenstate of H_s) $S(t=\infty) > S(0)$, i.e., in general, the pure state of the system becomes mixed in the course of time. This is a quantum result. Thus, the influence of the reservoir on the system *forces* one to consider the time evolution of mixed states. On the other hand, whereas in mixing flow the motion is time reversal invariant, it is not so in our approach. The reasons are obvious.

Furthermore, there is some doubt as to whether bound systems follow mixing flow.[11]

In the following section, it shall be proved that for any system and reservoir if $[H_R, H_I] = 0$, and if the system and the reservoir are initially uncorrelated, then the entropy of the system S is given by Eq. (5) will obey the relation $S(t) \geq S(0)$.

In Sec. III, $\rho_S(t)$ is evaluated approximately for the x-y model interacting with an Ising model with the assembly in a particular initial condition. The main purpose of this calculation is to show that the time it takes $\rho_S(t)$ to become diagonal is not unreasonably large even though there is no energy exchange between the system and the reservoir. In the case considered previously, Ising system interacting with an Ising reservoir, the time involved in the approach of $\rho_S(t)$ to a diagonal form turned out to be astronomically large. Thus, Sec. III supports the statement made in Ref. (15), that this undesirable feature was due to the Ising character of the model, and not to the basic mechanism we propose for thermal insulation. It is also shown in Sec. III that if the motion of the system is inverted at time T, then the Gibbs–Jaynes entropy of the system considered there at time τ later, $S^{MI}(T;\tau)$ is given by

$$S^{MI}(T;\tau) = S(T+\tau).$$

II. TIME EVOLUTION OF THE GIBBS-JAYNES ENTROPY OF A THERMALLY INSULATED SYSTEM, GENERAL

The proof given here is based on one given by Landau.[16]

Let the density matrix of a system interacting with a reservoir be ρ_S. Let the reservoir density matrix be ρ_R. The density matrix of the assembly (system + reservoir) shall be denoted by ρ_A. Clearly,[17]

$$\rho_S = \mathrm{Tr}_R \rho_A, \tag{8}$$

and

$$\rho_R = \mathrm{Tr}_s \rho_A, \tag{9}$$

where Tr_R (Tr_s) denotes the trace over reservoir (system) states only.

We shall consider arbitrary Hamiltonians for both the system and the reservoir, H_s and H_R, respectively. The reservoir Hamiltonian H_R will be such that

$$[H_R, H_I] = 0, \tag{10}$$

where H_I is the interaction Hamiltonian between system and reservoir.

We shall prove that if Eq. (10) holds, and if

$$\rho_A(t=0) = \rho_s(t=0)\rho_R(t=0) \tag{11}$$

with

$$\mathrm{Tr}_A \rho_A(0) = \mathrm{Tr}_s \rho_s(0) = \mathrm{Tr}_R \rho_R(0) = 1, \tag{12}$$

then

$$S_s(t) \geq S_s(t=0), \tag{13}$$

where S_s is the Gibbs–Jaynes entropy for the system defined by Eq. (5).

Equation (11) indicates that the system and reservoir are initially uncorrelated. A special case of Eq. (11) is provided by both the system and the reservoir being initially in a canonical distribution and brought together at $t = 0$, then one may, for example, remove a constraint in the system and follow its time evolution. Notice, however, that $\rho_s(0)$ and $\rho_R(0)$ are completely unspecified. Both the system and the reservoir might initially be in pure states, for example.

(a) It shall first be shown that if Eqs. (10)–(11) hold, then

$$\rho_s(t) = \sum_i \langle E_i^R | \rho_R(0) | E_i^R \rangle e^{-i(H_s + h_s^i)t} \rho_s(0) e^{+(H_s + h_s^i)t}, \tag{14}$$

where each symbol will become meaningful shortly. To start

$$\rho_s(t) = \sum_i \langle E_i^R | e^{-i(H_s+H_I)t} e^{-iH_R t} \rho_R(0) \times \rho_s(0) e^{+iH_R t} e^{+i(H_s+H_I)t} | E_i^R \rangle, \tag{15}$$

where $|E_i^R\rangle$ is an eigenstate of H_R. Since H_R commutes with H_I (Eq. 10) and with H_s, it drops out of the above equation. For the same reason, the state $|E_i^R\rangle$ may be chosen such that

$$H_I | E_i^R \rangle = h_s^i | E_i^R \rangle \tag{16}$$

where h_s^i is an operator acting only on the states of the system.

Notice that Eq. (14) gives $\rho_s(t)$ in terms of $\rho_s(0)$ with the evolution in time governed by a random static system Hamiltonian.

(b) Next, it will be shown that if Eq. (11) is satisfied, then

$$S_R(t) + S_S(t) \geq S_R(0) + S_S(0), \tag{17}$$

where

$$S_{(\,)} = -k\ \mathrm{Tr}\rho_{(\,)} \ln\rho_{(\,)}. \tag{18}$$

From Eq. (11) it follows that

$$\mathrm{Tr}_A \rho_A(0) \ln\rho_A(0) = \mathrm{Tr}_A \rho_S(0)\rho_R(0)[\ln\rho_S(0) + \ln\rho_R(0)], \tag{19}$$

since $[\rho_S, \rho_R] = 0$. Therefore,

$$S_A(0) = S_S(0) + S_R(0). \tag{20}$$

On the other hand,

$$S_A(t) = S_A(0), \tag{21}$$

and (see Ref. 18)

$$\mathrm{Tr}_A \rho_A(t) \ln\rho_A(t) - \mathrm{Tr}_A \rho_A(t) \ln[\rho_R(t)\rho_S(t)] \geq 0; \tag{22}$$

that is,

$$S_A(t) \leq S_S(t) + S_R(t). \tag{23}$$

Equations (20), (21) and (23) imply (17).

(c) Making use of the results obtained in parts (a)–(b) we shall now prove (13). Assume, temporarily that

$$\rho_R(0) = \rho_R(H_R), \tag{24}$$

then

$$\langle E_m^R|\rho_R(t)|E_l^R\rangle = \sum_i \langle E_m^R, E_i^s|e^{-iH_R t}e^{-i(H_s+H_I)t}\rho_R(0) \times \rho_s(0)e^{+i(H_s+H_I)t}\, e^{+iH_R t}|E_i^s, E_l^R\rangle, \tag{25}$$

since $[H_R,H_s] = [H_R,H_I] = 0$. Using Eq. (16), Eq. (25) becomes

$$\langle E_m^R|\rho_R(t)|E_l^R\rangle = e^{-i(E_m^R-E_l^R)t}\sum_i \langle E_m^R, E_i^s|e^{-i(H_s+h_s^m)t} \times \rho_R(0)\rho_s(0)e^{+i(H_s+h_s^l)t}|E_i^s, E_l^R\rangle, \tag{26}$$

but

$$[(H_s + h_s^m), \rho_R(0)] = [\rho_s(0), \rho_R(0)] = 0,$$

therefore, using (24) we have

$$\langle E_m^R|\rho_R(t)|E_l^R\rangle = \delta_{m,l}\langle E_m^R|\rho_R(0)|E_l^R\rangle \times \sum_i \langle E_i^s|e^{-i(H_s+h_s^m)t}\rho_s(0)e^{+i(H_s+h_s^l)t}|E_i^s\rangle. \tag{27}$$

Therefore,

$$\langle E_m^R|\rho_R^{(t)}|E_l^R\rangle = \delta_{m,l}\langle E_m^R|\rho_R(0)|E_l^R\rangle, \tag{28}$$

i.e.,

$$\rho_R(t) = \rho_R(0) \tag{29}$$

under the assumption (30). Thus,

$$S_R(t) = S_R(0)$$

which combined with Eq. (22) yields

$$S_s(t) \geq S_s(0).$$

But Eq. (14) implies that the value of $S_s(t)$ is the same whether (24) holds or not, therefore Eq. (13) holds for any $\rho_R(0)$ as long as Eqs. (10) and (11) hold. This completes the proof.

III. THE *XY* MODEL INTERACTING WITH AN ISING RESERVOIR

The model for the assembly (system + reservoir) consists of an x–y model (system) of N spins laid out on a circle, with its zero and first sites interacting with a one-dimensional Ising model (reservoir). The Hamiltonian for the assembly is given by

$$H = H_s + \left(\frac{B_0}{2}(\sigma_0^z + 1) + \frac{B_1}{2}(\sigma_1^z + 1)\right)\sum_{\mu=1}^{M} 2^{-\mu}\sigma_\mu^z + H_R, \tag{30}$$

where

$$H_s = \frac{J}{4}\sum_{i=0}^{N-1}(\sigma_i^x\sigma_{i+1}^x + \sigma_i^y\sigma_{i+1}^y), \tag{31}$$

$$H_R = \frac{1}{2}\sum_{\nu\mu} J'_{\nu\mu}\sigma_\nu^z\sigma_\mu^z, \tag{32}$$

and σ_ν^k is the kth component of the Pauli spin operator at the νth site. The Latin and Greek indices refer to system and reservoir sites, respectively. Clearly, there is no energy exchange between the system and reservoir, since

$$[H_R, H] = 0. \tag{33}$$

Furthermore,

$$[(H_s + H_I), H] = 0, \tag{34}$$

and

$$|\langle H_I\rangle| \leq |B_0| + |B_1|, \tag{35}$$

therefore, the energy of the reservoir $\langle H_s\rangle$ may fluctuate in time by a quantity of order 1, at most.

Performing the well-known Wigner–Jordan transformations

$$\sigma_i^x + i\sigma_i^y = 2C_i^\dagger : \left(\exp\, i\pi\sum_{j=0}^{i-1} c_j^+ c_j\right)$$

and

$$\sigma_i^x - i\sigma_i^y = 2C_i : \exp\left(-i\sum_{j=0}^{i-1} c_j^+ c_j\right)$$

(on the system but not on the reservoir), Eqs. (31) and (32) become

$$H_s = \tfrac{1}{2}J\sum_i (c_i^+ c_{i+1} + c_{i+1}^+ c_i), \tag{36}$$

and

$$H_I = (B_0 c_0^+ c_0 + B_1 c_1^+ c_1)\sum_{\mu=1}^{M} 2^{-\mu}\sigma_\mu^z. \tag{37}$$

It is convenient to perform a further transformation to obtain

$$H = \sum_k \Lambda_k \eta_k^+ \eta_k + H_R + \text{const} \tag{38}$$

with

$$[\eta_k, \eta_{k'}] = 0, [\eta_k, \eta_{k'}^+] = \delta_{kk'}. \tag{39}$$

Let

$$H_0 = B_0\sum_{\mu=1}^{M} 2^{-\mu}\sigma_\mu^z, \tag{40}$$

$$H_1 = B_1\sum_{\mu=1}^{M} 2^{-\mu}\sigma_\mu^z, \tag{41}$$

and

$$B_1 \ll B_0 \ll J \Longrightarrow H_1 \ll H_0 \ll J. \tag{42}$$

One may obtain Eq. (38) defining

$$\eta_k = \sum_i g_{ki} c_i, \tag{43}$$

and solving the equation

$$[\eta_k, H] = \Lambda_k \eta_k, \tag{44}$$

which, with H given by Eqs. (30), (32), (36), and (37), becomes

$$J(g_{k,j-1} + g_{k,j+1}) + 2\delta_{0,j}H_0 g_{k,0} + 2\delta_{ij}H_1 g_{k,1} = 2\Lambda g_{kj}. \tag{45}$$

A nontrivial solution of Eq. (45) exists if

$$\Lambda_k = J\cos k \tag{46}$$

for the values of k given by

$$\tan\left(\frac{Nk}{2}\right) = -\frac{H_0}{J\sin k}, \tag{47}$$

and

$$\tan\left(\frac{Nk}{2}\right) = -\frac{H_1 \sin k}{J} \tag{48}$$

under the conditions specified by Eq. (35), as can be shown after some tedious operations.

We arbitrarity choose the smallest $N/2$ positive solutions from Eq. (47) and the largest $N/2$ negative solutions from Eq. (48).

The solutions to Eq. (45) are then given by

$$g_{kj} = (2N)^{-1/2}\,(e^{ikj} - e^{-i\phi_k} e^{-ikj}), \quad \text{for } 1 \le j \le N, \tag{49}$$

and

$$g_{k0} \equiv g_{kN}, \tag{50}$$

where

$$e^{+i\phi_k} = [e^{-ikN} - 1 + (2H_1/J)e^{-ik}] \times [e^{ikN} - 1 + (2H_1/J)e^{-ik}]^{-1}. \tag{51}$$

The constant $(2N)^{-1/2}$ in Eq. (49) is chosen so that

$$\sum_j g_{kj}\tilde{g}_{kj} = 1. \tag{52}$$

The eigenvalue equation (45) implies

$$\sum_j g_{kj}\tilde{g}_{k'j} = 0 \quad \text{if } \Lambda_k \neq \Lambda_{k'} \tag{53}$$

From Eqs. (46), (47), and (48) it follows that $\Lambda_k \neq \Lambda_{k'}$ implies $k \neq k'$, and one can work out the case $k = k'$ to obtain

$$\sum_j g_{kj}\tilde{g}_{k'j} = \delta_{k,k'} \tag{54}$$

for the values of k and k' given by (47) and (48). Thus, Eq. (39) is satisfied.

To each distinct reservoir state (specified by each spin being either up or down) there corresponds a value of H_0 and a value of H_1, which, in turn, implies a distinct set of values for k (and, consequently, for Λ_k).

Thus, the Hamiltonian for the assembly is specified completely by Eqs. (32), (38), (39), and (46)-(48).

(a) Time evolution of the density matrix. An example

We are now in a position to consider a simple example of an initial condition for the assembly and follow the time evolution of the density matrix for the system.

The time evolution of the system will now be followed from the assembly's density matrix initially given by

$$\rho_A(0) = \rho_s(0)\rho_R(0), \tag{55}$$

where $\rho_s(0)$ and $\rho_R(0)$ are the initial system and reservoir density matrices, respectively. The following simple initial condition will be considered

$$\rho_s(0) = 2^{-N} + 2^{-N}\sum_j h_j(c_j^+ c_j - \tfrac{1}{2}), \tag{56}$$

and

$$\rho_R(0) = 2^{-M}. \tag{57}$$

$\rho_s(0)$ is nondiagonal in the representation of H_0 and satisfies $\mathrm{Tr}_s\rho_s(0) = 1$ (Tr_s stands for the trace over the states of the system only). Furthermore, $h_j \ll 1$ for all j. Thus, the reservoir is initially at an infinite temperature, while the system is slightly off-equilibrium. For example, $4\langle S_j^z(t=0)\rangle = h_j\hbar$.

The system and the reservoir are obviously uncorrelated at $t = 0$. The reservoir is at infinite temperature, whereas the system itself may be thought of as a collection of noninteracting spins (for $t < 0$) in thermal equilibrium under a weak external field varying with position. One would thus, have Eqs. (56) and (57) at $t = 0$. This interpretation would imply that the external field is switched off at $t = 0$ and, simultaneously, the Hamiltonian given by Eqs. (30)-(33) is turned on.

We are interested in evaluating

$$\rho_s(t) = \sum_{\{m_\mu\}} \langle\{m_\mu\}|\rho_A(t)|\{m_\mu\}\rangle, \tag{58}$$

where $|\{m_\mu\}\rangle$ is a reservoir state with the spin at the νth site up [down if $m_\nu = +1\ (-1)$]. Taking the $B_0 \to 0$ limit after the $B_1 \to 0$ limit is taken, it follows from Eqs. (47)-(51) that the set of g_{Qj} become

$$g_{Qj} = 2^{1/2}N^{-1/2}\cos(Qj) \quad \text{for } 0 < Q \le \pi, \tag{59}$$

$$g_{Qj} = i2^{1/2}N^{-1/2}\sin(Qj) \quad \text{for } \pi \le Q < 0, \tag{60}$$

where Q and P are given by $2\pi n/N$, with $n = -(N/2'), -(N/2)+1, \ldots, -1, 1, \ldots, (N/2)$.

Substituting Eqs. (59), (60), and (43) into Eq. (58), the following equation obtains:

$$\rho_s(t) = \sum_{\{m_\mu\}} 2^{-M}\langle\{m_\mu'\}|e^{-iH\hbar^{-1}t}\Big(\rho_s^D(0) + \sum_{Q,P} 2^{-N}G(Q,P)\eta_Q^+\eta_P - 2^{-(N+1)}\sum_j h_j\Big)e^{+iH\hbar^{-1}}|\{m_\mu\}\rangle, \tag{61}$$

where $G(Q,P) = \sum_j g_{Qj}\tilde{g}_{Pj}h_j$. It is shown in the Appendix that a negligible error is introduced if one uses

$$\eta_Q(t) = \sum_k \Delta_{Q,K}\eta_k(0)e^{-i\Lambda_k\hbar^{-1}t}, \tag{62}$$

where $\Delta_{Q,K} = 1$ if $|k - Q| < \pi/N$ and vanishes otherwise, k is given by Eqs. (47) and (48), $Q = 2\pi n/N$. Substituting Eq. (62) and its Hermitian conjugate into Eq. (61), one obtains

$$\rho_s(t) = \sum_{\{m_\mu\}} 2^{-M}\langle\{m_\mu\}|\sum_{Q,P} e^{-i\hbar^{-1}tJ(\cos Q - \cos P)}2^{-N}G(Q,P)\eta_Q^+\eta_P \exp\left[-\left(i\hbar^{-1}t\sum_k \Lambda_k(H_0,H_1)(\Delta_{kQ} - \Delta_{k'Q})\right)\right]|\{m_\mu\}\rangle + \left(\rho_s^D(0) - 2^{-(N+1)}\sum_j h_j\right). \tag{63}$$

Now, it follows from Eqs. (46)-(48), and $B_1 \ll B_0 \ll J$, that to a very good approximation wehave

$$\Lambda_{k(H_0,H_1)} = \sum_Q \Delta_{Q,k}\Big(J\cos Q + \frac{2}{N}(B_0 + B_1\cos^2 Q) \times \sum_{\mu=1}^{M} 2^{-\mu}m_\mu\Big), \quad \text{for } k > 0 \tag{64}$$

and

$$\Lambda_{k(H_0,H_1)} = \sum_Q \Delta_{Q,K}\Big[J\cos Q + \frac{2B_1}{N}\sin^2 Q \times \sum_{\mu=1}^{M} 2^{-\mu} m_\mu\Big], \quad \text{for } k<0, \tag{65}$$

except for a very small region for $k>0$ (which we shall neglect) where $B_0 \ll J(\sin k)^{-1}$ is not satisfied. Substituting Eqs. (64) and (65) into Eq. (63), *taking the* $M\to\infty$ *limit*, and using the simple identity[19] $x^{-1}\sin x = \Pi_{\mu=1}^{\infty}\cos(2^{-\mu}x)$, one obtains

$$\rho_s(t) = 2^{-N} - 2^{-(N+1)}\sum_j h_j + 2^{-N}\sum_{Q,P} G(Q,P)\eta_Q^+\eta_P W(Q,P,t) \times e^{-iJ\hbar^{-1}t}(\cos Q - \cos P), \tag{66}$$

where

$$W(Q,P,t) = [2N^{-1}F(Q,P)\hbar^{-1}t]^{-1}\sin[2N^{-1}F(Q,P)\hbar^{-1}t],$$

and

$$F(Q,P) = \theta(Q)\theta(P)B_1(\cos^2 Q - \cos^2 P) + [\theta(Q)\theta(-P) - \theta(-Q)\theta(P)][B_0 + B_1(\cos^2 Q - \cos^2 P)] + \theta(Q)\theta(-P)B_1(\sin^2 Q - \sin^2 P). \tag{67}$$

Clearly, all-off diagonal elements of $\rho_s(t)$ vanish in the $t\to\infty$ limit, in contradistinction to the case where the system of interest is dynamically isolated, where

$$\langle E_i|\rho_s(t)|E_j\rangle = \exp[i(E_j - E_i)t].$$

It should be noticed that the off-diagonal terms of $\rho_s(t)$ decrease significantly in magnitude in a time $\tau \sim (N\hbar/J)(J/B_0)$ or $\tau \sim (N\hbar/J)J/B_1)$, which may be interpreted as the time it takes a typical excitation in the system to traverse it times (J/B_1) or (J/B_0). In the case $B_1\to 0$, $B_0\to 0$, one recovers the isolated x–y model result, i.e., $\tau=\infty$.

(b) The Gibbs-Jaynes entropy

The Gibbs–Jaynes entropy for the system,

$$S(t) = -k_B\,\mathrm{Tr}_s[\rho_s(t)\ln\rho_s(t)], \tag{68}$$

fulfills

$$S(t=\infty) > S(t=0),$$

since $\rho_s(t)$ becomes diagonal in the representation of H_s as $t\to\infty$, and it follows from Peierls' theorem for convex functions that

$$\mathrm{Tr}\rho\ln\rho \geq \mathrm{Tr}\rho_{ii}\ln\rho_{ii},$$

the equality sign holding only if $\langle i|\rho|j\rangle = \delta_{ij}\rho_{ii}$.

Thus, the case under consideration provides an example of actual entropy increase, and in a reasonable time τ.

To be sure that $S(t=\infty) - S(t=0)$ is not negligible, we shall obtain an expression for $S(t)$ to second order in $h_j's$. It is a bit tedious, but straightforward, to substitute Eq. (66) into (68) to obtain

$$S(t) = Nk_B\ln 2 - (k_{B/8})\sum_Q [G(Q,Q)]^2 - (k_{B/8})\sum_{P,Q}{}' [G(Q,P)]^2[W(Q,P,t)]^2, \tag{69}$$

where the prime on the summation sign means $Q=P$ not included, whence

$$S(t=\infty) - S(t=0) = (k_{B/8})\sum_{P,Q}{}' [G(Q,P)]^2 \tag{70}$$

or equivalently,

$$S(t=\infty) - S(t=0) = (k_{B/8})\sum_i h_i^2, \tag{71}$$

which is in general certainly not negligible. The second term in Eq. (69) is clearly time independent, and is due to the fact that the total spin $\sum\sigma_i^z$ is a constant of the motion, as may be easily checked.

(c) Motion inversion

We shall consider here an arbitrary initial state for the assembly (xy model + Ising reservoir). The assembly is allowed to evolve in time until $t=T$. Let the system density matrix at this time be $\rho_s(T)$. At $t=T$ the motion of the system (not the whole assembly) is inverted, i.e., all the spins in the system are "turned around," but not the spins in the reservoir. Let the system density matrix at time τ after the motion inversion takes place be $\rho^{MI}(T;\tau)$. Now, if the system were *dynamically* isolated (instead of *thermally* insulated as it is here) we would have

$$\rho^{MI}(T;T) = \rho^{MI}(0,0), \tag{72}$$

i.e., the motion of the system would be time reversal invariant. It will be shown next, that this is not true in the case under consideration. The assembly's density matrix immediately after T, $\rho_A^{MI}(T)$, is given by

$$\langle R,S|\rho_A^{MI}(T)|S',R'\rangle = \langle R,S'|U\rho_A(T)U^{-1}|S,R'\rangle, \tag{73}$$

where $|S,R\rangle$ denotes the system in state S and the reservoir in state R, $U=\Pi_i\sigma_i^Y$ (the representation of S_j^z for every j is assumed), and the product Π_j is over the whole system (but *not* the reservoir). One may check the validity of Eq. (73) verifying that any spin belonging to the system is inverted while the reservoir is unaffected by the motion inversion operation for any $\rho_A(\tau)$. It follows, without difficulty, from Eq. (73) that the density matrix for the system at time τ after the motion inversion operation takes place is given by

$$\rho_S^{MI}(T;\tau) = \sum_R \langle R|e^{-iH\tau}U\tilde{\rho}_A(T)U^{-1}e^{+iH\tau}|R\rangle, \tag{74}$$

where the tilde denotes complex conjugation. Using $\tilde{H}=H$ and $Uf(H_s+H_R+H_I)U^{-1} = f(H_s+H_R-H_I)$ for any function f, it follows that

$$\rho_S^{MI}(T;\tau) = \sum_R \langle R|e^{-i(H_s+H_R+H_I)\tau}e^{+i(H_s+H_R-H_I)T} \times U\tilde{\rho}_A(0)U^{-1}e^{-i(H_s+H_R-H_I)T}e^{+i(H_s+H_R+H_I)\tau}|R\rangle. \tag{75}$$

Now, since $[H_R,H_I] = [H_R,H_S] = 0, H_R$ disappears from the above equation. Using Eqs. (9), (62), that $\Lambda_k - \Lambda_Q$ is odd in $(k-Q)$ (neglecting errors of order $1/N$) for $|Q-k| < \pi/N$, Eq. (75) becomes

$$\rho_S^{MI}(T;\tau)e^{+iH_s 2T}\sum_R \langle R|\exp\Big[-i\sum_Q\Big(\sum_k \Delta_{k,Q}\Lambda_k\Big)\eta_Q^+\eta_Q(T+\tau) \times U\tilde{\rho}_A(0)U^{-1}\exp\Big[i\sum_Q\Big(\sum_k \Delta_{Q,k}\Lambda_k\Big) \times \eta_Q^+\eta_Q(T+\tau)|R\rangle e^{+iH_s 2T}. \tag{76}$$

The operators U and U^{-1} may now be omitted from the

above equation. So,

$$\rho_s^{MI}(T;\tau) = e^{+iH_s 2T}\sum_R \langle R|\exp[-i\sum_Q\Big(\sum_k \Delta_{k,Q}\Lambda_k\Big)$$
$$\times\eta_Q^+\eta_Q(-T-\tau)\rho_A(0)\exp[i\sum_Q\Big(\sum_k \Delta_{k,Q}\Lambda_k\Big)$$
$$\times\eta_Q^+\eta_Q(-T-\tau)|R\rangle e^{-iH_s 2T} \tag{77}$$

and consequently

$$\rho_s^{MI}(T;\tau) = e^{iH_s 2T}\rho_s(-T-\tau)e^{-iH_s 2T}. \tag{78}$$

Clearly, then,

$$S_s^{MI}(T;\tau) = S_s(-T-\tau), \tag{79}$$

since S is a real quantity. It is not difficult to see that

$$S_s(-t) = S_s(t); \tag{80}$$

therefore,

$$S_s^{MI}(T;\tau) = S_s(T+\tau). \tag{81}$$

Equation (80) implies that $S(t)$ *decreases* for $t<0$. Since $S(t)$ increases for $t>0$. There is no contradiction here. It was shown in Sec. II that $S(t)\geq S(0)$ *if* $\rho_A(0) = \rho_s(0)\rho_R(0)$. Consequently, for $t<0$, S may decrease, as it may during any interval of time if the system and the reservoir are correlated to start with. On the other hand, Eqs. (69), (81) show that the entropy does not in general return to its initial value if the motion is inverted some time T after the initial condition $\rho_A = \rho_s\cdot\rho_R$ is fulfilled.

APPENDIX

It is shown here that, to a good approximation

$$\eta_Q(t) = \eta_Q(0)\sum_k \Delta_{k,Q}e^{-i\Lambda_k t}, \tag{A1}$$

where

$$\Delta_{Q,k} = \begin{cases} 1 & \text{if } |k-Q|<\pi/N, \\ 0 & \text{otherwise,}\end{cases} \tag{A2}$$

$QN = 2\pi n$, and k is given by Eqs. (47) and (48).

We start with

$$i\dot\eta_Q(t) = [\eta_Q(t), H], \tag{A3}$$

which upon substitution of Eq. (38) becomes

$$i\dot\eta_Q(t) = \sum_k[\eta_Q(t),\eta_k^+(t)]\Lambda_k\eta_k(t). \tag{A4}$$

Now, using Eq. (43) one obtains

$$[\eta_Q(t),\eta_k^+(t)] = \sum_j g_{Qj}\tilde g_{kj}, \tag{A5}$$

and by definition, g_{Qj} satisfies

$$(J/2)(g_{Q,j}+g_{Q,j-1}) = \Lambda_k g_{Q,j}, \tag{A6}$$

whereas g_{kj} satisfies Eq. (45). If one multiplies (A6) by g_{kj} and sums on j, similarly multiplies (45) by g_{Qj} and sums on j, and finally subtracts the two equations, one obtains

$$(\Lambda_k-\Lambda_Q)\sum_j \tilde g_{Qj}g_{kj} = H_1 g_{k1} + H_0 g_{k0}\tilde g_{Q0}. \tag{A7}$$

Use of Eqs. (49), (59), and (60) yields

$$\sum_j g_{Qj}\tilde g_{kj} = -\frac{1}{N}\frac{H_0(e^{ikN}-e^{i\phi k}e^{-ikN})}{\Lambda_Q-\Lambda_k} \quad \text{for } Q>0, \tag{A8}$$

and

$$\sum_j g_{Qj}\tilde g_{kj} = -\frac{i}{N}H_1\sin Q\frac{(e^{-ik}-e^{i\phi k}e^{+ik})}{\Lambda_Q-\Lambda_{-k}} \text{ for } Q<0. \tag{A9}$$

With the help of Eqs. (46), (48), and (51) the above equations become

$$\sum_j g_{Qj}\tilde g_{kj} = \Delta_{Q,k} + \Big(\frac{H_0}{J\sin k}\Big) \quad \text{for } Q>0, \tag{A10}$$

and

$$\sum_j g_{Qj}\tilde g_{kj} = \Delta_{Q,k} + 0\Big(\frac{H_1}{\sin k}\Big) \quad \text{for } Q<0. \tag{A11}$$

Neglecting the small region where $B_0/J\sin k \ll 1$ is not fulfilled, we have from Eqs. (A5), (A10), and (A11)

$$\eta_Q(t) = \sum \Delta_{Q,k}\eta_k(t). \tag{A12}$$

Furthermore,

$$\eta_k(t) = \eta_k(0)e^{-i\Lambda_k t}. \tag{A13}$$

These last two equations yield Eq. (A1).

[1] A. Messiah, *Quantum Mechanics* (Wiley, New York, 1961), p. 331; A. S. Davydov, *Quantum Mechanics* (Pergamon, Oxford, 1965), p. 38; K. Huang, *Statistical Mechanics* (Wiley, New York, 1963), p. 187.

[2] J. M. Blatt, Prog. Theor. Phys. **22**, 745 (1959).

[3] J. J. Sakurai, *Invariance Principles and Elementary Particles* (Princeton, U. P., Princeton, N. J., 1964), p. 79.

[4] E. L. Hahn, Phys. Rev. **80**, 580 (1950); Phys. Today **6**, 4 (1953); S. Hartman and E. L. Hahn, Phys. Rev. **128**, 2042 (1962).

[5] W. K. Rhim, A. Pines, and J. S. Waugh, Phys. Rev. Lett. **25**, 218 (1970).

[6] E. T. Jaynes, Phys. Rev. **106**, 620 (1957); Phys. Rev. **108**, 171 (1957); and in *Statistical Physics*, edited by G. E. Uhlenbeck, N. Rosenzweig, A. F. G. Siegert, E. T. Jaynes, and S. Fujita (Benjamin, New York, 1963), p. 182.

[7] The derivation of the master equation with coarse graining may be found in: N. G. van Kampen, in *Fundamental Problems in Statistical Mechanics*, edited by E. G. D. Cohen (North-Holland, Amsterdam, 1962), p. 173; O. Penrose, *Foundations of Statistical Mechanics* (Pergamon, Oxford, 1970).

[8] For a simplified derivation and further references, see R. Zwanzig, *Lectures in Theoretical Physics*, Vol. III, edited by W. E. Brittin, D. W. Downs, and J. Downs (Interscience, New York, 1961), p. 106.

[9] R. F. Fox, J. Math. Phys. **13**, 1196 (1972); J. Math. Phys. **13**, 1726 (1972); J. Math. Phys. **14**, 20 (1973).

[10] J. Lebowitz and O. Penrose, Phys. Today **26**, 23 (1973).

[11] J. Ford, Adv. Chem. Phys. (to be published).

[12] J. Lebowitz, *Statistical Mechanics*, edited by S. A. Rice, K. F. Freed, and J. C. Light (University of Chicago Press, Chicago, 1972), p. 41.

[13] Ya. Sinai, *Statistical Mechanics Foundations and Applications*, edited by T. A. Bak (Benjamin, New York, 1967), p. 559.

[14] P. G. Bergmann and J. L. Lebowitz, Phys. Rev. **99**, 578 (1955); E. P. Gross and J. L. Lebowitz, Phys. Rev. **104**, 1528 (1956); J. L. Lebowitz and H. L. Frisch, Phys. Rev. **107**, 917 (1957); J. L. Lebowitz and P. G. Bergmann, Ann. Phys. (N.Y.) **1**, 1 (1959); M. A. Huerta and H. S. Robertson, J. Stat. Phys. **1**, 393 (1969); H. S. Robertson and M. A. Huerta, Phys. Rev. Lett. 23, 825 (1969); Am. J. Phys. **38**, 619 (1970); G. S. Agarwal, Phys. Rev. **3**, 828 (1971); G. G. Emch and G. L. Sewell, J. Math. Phys. **9**, 946 (1968); B. Roberston and W. C. Mitchell, J. Math. Phys. **12**, 563 (1971); G. G. Emch and C. Radim, J. Math. Phys. **12**, 2043 (1971).

[15] J. F. Fernández, Phys. Lett. A **35**, 15 (1971).

[16] L. D. Landau, *Collected Papers of L. D. Landau*, edited by D. ter Haar (Pergamon, Oxford, 1965), p. 70.

[17] V. Fano, Rev. Mod. Phys. **29**, 74 (1957).

[18] S. Okubo and A. Isihara, J. Math. Phys. **12**, 2498 (1971).

[19] S. P. Heims, Am. J. Phys. **33**, 722 (1965); M. Kac, *Statistical Independence in Probability, Analysis and Number Theory* (Mathematical Association of America, New York, 1959), p. 1.

Integral equations for scattering in potentials of infinite range

D. F. Freeman and J. Nuttall*

Physics Department, University of Western Ontario, London, Ontario, Canada
(Received 8 June 1973)

We show how, by finding an approximation to the exact Green's function, to obtain integral equations with compact kernels for two scattering problems involving potentials of infinite range. They are one particle scattering in a Coulomb potential for a given partial wave and one particle scattering in three dimensions in a potential having r^{-2} behavior at infinity. In the second case, we solve only an inhomogeneous version of Schrödinger's equation.

1. INTRODUCTION

There has been a great deal of activity in the area of nonrelativistic quantum mechanical three-particle scattering since the demonstration by Faddeev[1] that the Watson integral equations for multiple scattering could be treated on a rigorous mathematical basis. Actually, relatively little use has been made in numerical calculations of Faddeev's result that some power of the kernel of the integral equation is compact in a suitable Banach space. Perhaps the main effect has been to give people confidence that the problem is not really as difficult as had been imagined.

Faddeev's work does not apply to certain potentials of infinite range, including the Coulomb potential and potentials behaving like r^{-2} at infinity. Our aim is to derive well-behaved integral equations for these potentials. We are not particularly optimistic that the use of such equations will be the best way to obtain numerical results in three-particle scattering problems, but perhaps the existence of the equations would help to build confidence, as in the short-range case. In addition, they should provide a firm basis to discuss such questions as the ionization threshold behavior.

Before considering the three-particle problem, we must first understand the one-particle problem. As far as we know, there is no satisfactory integral equation for the scattering of a particle in a Coulomb potential, even in partial waves, and none for scattering in a r^{-2} potential in three dimensions. This paper begins the task outlined above by finding integral equations for two problems. The first is one-particle Coulomb scattering in a given partial wave and the second is one-particle scattering in three dimensions in a potential behaving like r^{-2} at infinity. Here our results are somewhat restricted, as will be explained below.

Our method involves finding an apporximation $\bar{G}$ to the exact Green's function G which is more accurate at large distances than the commonly used G_0, the resolvent of the kinetic energy operator. In the cases discussed in this paper, the straight-line eikonal approximation provides the basis for a satisfactory form of $\bar{G}$, and perhaps this result will generalize to other problems involving long-range interactions.

2. PARTIAL WAVE COULOMB SCATTERING

To illustrate our approach, we first discuss the problem of one particle scattering in a Coulomb potential in a given partial wave, which we take for simplicity to be the S wave. We could easily remove this restriction and also add a short-range potential.

The long-range nature of the potential causes difficulties in both initial and final states. In many three-particle problems, such as scattering of an electron from a neutral atom, the initial state is easy to deal with, and it is only the final state that causes trouble. In a one particle problem, it is difficult to achieve this situation. To come as near as possible, we study an inhomogeneous version of Schrödinger's equation for the wavefunction ψ,

$$(k^2 - H)\psi = g, \tag{1}$$

where g is a function that falls off rapidly at large distances and ψ contains only outgoing waves at infinity. The e–H scattering problem could be stated in this way.

In the present case we have

$$H = -\frac{d^2}{dr^2} + \frac{c}{r} = -\frac{d^2}{dr^2} + V(r),$$

$$\psi(r) = 0 \text{ at } r = 0.$$

The Lippmann–Schwinger equation for ψ is obtained from (1) by operating with $G_0(r, r')$, where

$$G_0(r, r') = \sin kr_< \, e^{ikr_>}.$$

The kernel of the integral equation, $G_0 V(r')$, is not L^2, nor, as far as we can see, is it compact. The situation is not improved by explicitly including the leading term in the expansion of ψ for large r.

We consider $\bar{G}(r, r')$, chosen to represent more accurately the behavior of the exact G for large r, r'. We take

$$\bar{G}(r, r') = \sin[\phi(r_<)] \exp[i\phi(r_>)],$$

where

$$\phi(r) = kr + \eta \ln(r + a) - \eta \ln a \text{ and } \eta = -c/2k.$$

We have omitted the phase shift because this is related to the short-range form of $V(r)$. We have used information that could have been found from the JWKB approximation, without having the exact Coulomb wavefunctions.

Applying $\bar{G}$ to (1) and integrating by parts twice leads to the equation

$$\int_0^\infty dr' (k^2 - H')\bar{G}(r, r')\psi(r') + \phi'(r)\{i \sin\phi(r) \exp[\phi(r)] - \cos\phi(r) \exp[\phi(r)]\}\psi(r) = \int_0^\infty dr' \bar{G}(r, r')g(r'),$$

which may be written as

$$\psi(r) + \int_0^\infty dr' K(r, r')\psi(r') = f^{-1}(r) \times \int_0^\infty dr' \bar{G}(r, r')g(r')$$

with

$$f(r) = -[k + \eta/(r + a)]$$

and

$$K(r,r') = f^{-1}(r)(k^2 - H')\bar{G}(r,r')$$
$$= \begin{cases} -f^{-1}(r)\sin\phi(r)\exp[i\phi(r)] \\ \times \{[ac/r'(r'+a)] + (i\eta + \eta^2)/(r'+a)^2\}, r' > r, \\ -f^{-1}(r)\exp[i\phi(r)]\Big(\sin\phi(r')\{[ac/r'(r'+a)] \\ + \eta^2/(r'+a)^2\} + \cos\phi(r')\dfrac{\eta}{(r'+a)^2}\Big), \quad r' < r. \end{cases}$$

Let us choose a so that

$$|\eta/a| < k,$$

in which case $f(r)$ will not vanish. It may be seen that the kernel of the integral equation $K(r,r')$ sends functions dominated as $r \to \infty$ by r^α, $0 < \alpha < 1$, into functions bounded by a constant at ∞. From this we deduce that $K(r,r')$ represents a compact operator in the Banach space[2] of functions continuous on $0 < r < \infty$ with norm

$$\|f\| = \sup_{0<r<\infty} |f(r)(1+r)^{-\alpha}|.$$

Alternatively, it would be possible to obtain an equation with an L^2 kernel by writing

$$\psi(r) = t\exp[i\phi(r)](1 - e^{-r}) + \theta(r)$$

and deducing an equation for $\theta(r)$ with the help of the relation

$$t + k^{-1}\int_0^\infty dr'\left[\left(\frac{ac}{r'(r'+a)} + \frac{\eta^2}{(r'+a)^2}\right)\sin\phi(r') + \frac{\eta}{(r'+a)^2}\cos\phi(r')\right]\psi(r')$$
$$= -k^{-1}\int_0^\infty dr' g(r')\sin\phi(r').$$

We can use the same formalism to discuss the homogeneous form of Schrödinger's equation. We look for a solution of

$$(E - H)\chi = 0$$

of the form

$$\chi = \psi + \sin\phi(r)$$

with ψ containing only outgoing waves at infinity. Then ψ satisfies (1) with g given by

$$g(r) = \sin\phi(r)\left(\frac{ac}{r(r+a)} + \frac{\eta^2}{(r+a)^2}\right) + \frac{\eta}{(r+a)^2} \times \cos\phi(r),$$

and the analysis described above applies.

3. THE CASE OF THE r^{-2} POTENTIAL

We now study the case of one particle scattering in a potential with r^{-2} behavior at infinity, but do not make a partial wave expansion. As an example, we take the potential $V(r) = c(r^2 + b^2)^{-1}$. Again the kernel of the Lippmann–Schwinger equation is not L^2, and we do not know how to make it compact. To derive an integral equation with a compact kernel, we find a suitable approximation $\bar{G}$ to the exact Green's function $G(\mathbf{r},\mathbf{r}')$ which satisfies

$$(k^2 - H')G(\mathbf{r},\mathbf{r}') = -4\pi\delta(\mathbf{r}' - \mathbf{r}).$$

The region of most importance is that in which both $|\mathbf{r}|$ and $|\mathbf{r}'|$ are large. In almost all of this region, we can find a satisfactory $\bar{G}$ by using the eikonal approximation with trajectories which are straight lines diverging from the point $\mathbf{r}' = \mathbf{r}$. This leads, after approximating $[k^2 - V(r)]^{1/2}$ by $k - C/2kr^2$, to

$$\bar{G}_1(\mathbf{r},\mathbf{r}') = (e^{ik\rho}/\rho)e^{i\Phi_1}, \tag{2}$$

where

$$\Phi_1 = -\frac{c}{2k}\int_0^1 dt(\mathbf{r} + t\boldsymbol{\rho})^{-2}$$
$$= -\frac{c}{2k}\frac{\theta}{p}. \tag{3}$$

Here, we have used $\boldsymbol{\rho} = \mathbf{r}' - \mathbf{r}$, $\cos\theta = \hat{\mathbf{r}}'\cdot\hat{\mathbf{r}}$ and $p = (rr'/\rho)\sin\theta$.

We cannot expect this approximation to G to be valid near $\theta = \pi$, since the trajectory must pass near the origin where the potential is not small. To avoid singularities in this region, we modify (3) to read

$$\Phi_2 = -(c/2k)\,\theta(p^2 + l^{1/2} + 1)^{-1/2}, \tag{4}$$

where

$$l = rr'/\rho.$$

The expression (4) is close to (3) for large l except near $\theta = \pi$.

A further modification of (2) is needed to improve the behavior near $r = 0$. Take twice differentiable functions $h_1(\mathbf{r}), h_2(\mathbf{r})$ satisfying

$$h_1 + h_2 = 1,$$
$$h_1 = 0, \quad r > 1,$$
$$h_2 = 0, \quad r < \tfrac{1}{2}.$$

Then our final approximation $\bar{G}$ is

$$\bar{G}(\mathbf{r},\mathbf{r}') = (e^{ik\rho}/\rho)[h_1(\mathbf{r}) + h_2(\mathbf{r})e^{i\Phi_2}]. \tag{5}$$

We are again interested in solving (1), where in this case

$$H = -\nabla^2 + c(r^2 + b^2)^{-1}$$

and

$$\psi(\mathbf{r}) \underset{r\to\infty}{\sim} f(\hat{\mathbf{r}})e^{ikr}/r.$$

Take (1) with $\mathbf{r}'$ as the independent variable, multiply by (5) and integrate over the region between two spheres centered at $\mathbf{r}' = \mathbf{r}$, one small of radius R_1 and the other large of radius R_2. Use Green's theorem and let $R_1 \to 0, R_2 \to \infty$; we obtain

$$-4\pi\psi(\mathbf{r}) + \int d\mathbf{r}' K(\mathbf{r},\mathbf{r}')\psi(\mathbf{r}') = \int d\mathbf{r}'\bar{G}(\mathbf{r},\mathbf{r}')g(\mathbf{r}') \tag{6}$$

and

$$K(\mathbf{r},\mathbf{r}') = (k^2 - H')\bar{G}(\mathbf{r},\mathbf{r}').$$

Our contention is that the kernel $K(\mathbf{r},\mathbf{r}')$ of the integral equation (6) represents a compact operator in the

Banach space[2] of continuous functions $\psi(\mathbf{r})$ with norm given by

$$\|\psi\| = \sup_{r} |\psi(\mathbf{r})(1+r)^{\alpha}|.$$

The parameter α must be chosen to satisfy $0 < \alpha < 1$. An outline of the proof of this statement follows.

First we evaluate $K(\mathbf{r},\mathbf{r}')$ and find, using $d^2 = p^2 + l^{1/2} + 1$,

$$\begin{aligned} K(\mathbf{r},\mathbf{r}') = {} & (cb^2r'^{-2}(r'^2+b^2)^{-1} + cr'^{-2}d^{-1}(p+d)^{-1} \\ & \times (l^{1/2}+1) + c\theta l^{1/2}\mathbf{r}\cdot\mathbf{r}'\rho^{-1}d^{-3}r'^{-2} - (ic/2k) \\ & \times \{d^{-1}r'^{-2}\rho^{-1}p^{-1}(\mathbf{r}\cdot\mathbf{r}' - 2p^2) + d^{-3}[\theta\rho^{-2}(3p^2 \\ & - 2r^2) + \tfrac{1}{8}\theta r\rho^{-3}r'^{-1}l^{-1/2}(3r^2+4p^2) \\ & + 2p\rho^{-1}r'^{-2}(p^2 - \mathbf{r}\cdot\mathbf{r}')] \\ & + 3\theta d^{-5}[p^2\rho^{-2}(r^2-p^2) + \tfrac{1}{16}r^3\rho^{-3}r'^{-1} \\ & - \tfrac{1}{4}rp^2\rho^{-3}r'^{-1}l^{-1/2}\mathbf{r}\cdot\rho]\} + (c^2/4k^2) \\ & \times \{d^{-2}r'^{-2} + \theta d^{-4}[2p\rho^{-1}r'^{-2}(p^2 - \mathbf{r}\cdot\mathbf{r}') \\ & + \tfrac{1}{2}rp\rho^{-2}r'^{-1}l^{-1/2}] + \theta^2 d^{-6}[p^2\rho^{-2}(r^2-p^2) \\ & + \tfrac{1}{16}r^3\rho^{-3}r'^{-1} - \tfrac{1}{2}rp^2\rho^{-3}r'^{-1}l^{-1/2}\mathbf{r}\cdot\rho]\}) \\ & \times h_2(\mathbf{r})e^{i(kp+\phi_2)}/p - V(r')h_1(\mathbf{r})e^{ik\rho}/\rho. \end{aligned} \tag{7}$$

With this we show that, if $f(\mathbf{r})$ is a continuous function with $\|f\| = 1$, then $F(\mathbf{r})$ given by

$$F(\mathbf{r}) = \int d\mathbf{r}' K(\mathbf{r},\mathbf{r}')f(\mathbf{r}') \tag{8}$$

satisfies

$$\sup_{\mathbf{r}} |F(\mathbf{r})(1+r)^{\beta}| < C, \tag{9}$$

where $\alpha < \beta < 1$ and C is independent of f. The estimates necessary to show this are achieved routinely by considering separately three regions that contribute to (8), namely

(i) $r' \geq 2r$,
(ii) $r' \leq 2r$, $\rho \geq \frac{1}{2}r$,
(iii) $0 \leq \rho \leq \frac{1}{2}r$.

Some further details are given in the Appendix.

We also need the equicontinuity of all functions $F(\mathbf{r})$, in any region $|\mathbf{r}| < R$, which is demonstrated by using

$$|F(\mathbf{r}_1) - F(\mathbf{r}_2)| < \int d\mathbf{r}' |K(\mathbf{r}_1,\mathbf{r}') - K(\mathbf{r}_2,\mathbf{r}')|(1+r')^{\alpha}$$

and showing that the right-hand side, which is independent of F, approaches zero as $|\mathbf{r}_1 - \mathbf{r}_2| \to 0$ for any fixed $\mathbf{r}_2$.

With these two results, the compactness of K follows immediately by standard techniques, much as in the proof of Faddeev.[1]

4. DISCUSSION

There is a chance that the same sort of approach, with suitable modifications, will apply to Coulomb scattering in three dimensions and to three-particle problems involving long-range potentials. Our lack of knowledge of the form of G in what might be called the forward direction did not impede the derivation of a good integral equation. However, a similar ignorance about the way in which an incident plane wave is modified in this direction meant that we were forced to consider only an inhomogeneous form of Schrödinger's equation. To solve the homogeneous form of problems such as that discussed in Sec. 3, it will be necessary to either learn more about the wavefunction in the forward direction or discover an improved form of the integral equation.

APPENDIX

To demonstrate (9), we need only consider the coefficient of $h_2(\mathbf{r})$ in (7). It is straightforward to show that we can find a constant such that

$$\int dr' d^{-a}\theta^b p^c l^e \rho^f r'^j (1+r')^{-\alpha} < C(1+r)^{-\beta} + \epsilon, \tag{A1}$$

where ϵ is any positive number and β is the minimum value of

$$\begin{aligned} & \alpha + a - c - e - f - j - 3 - A, \\ & -f - \max(0, 3 + j + c + e - a - \alpha + A), \\ & \alpha + a - c - e - f - j - 3, \end{aligned}$$

provided that

$$\begin{aligned} & 2 + j + f - \alpha < 1, \\ & 2 + j + e + c > 1, \\ & 2 + b - e + f + a/4 > 1, \\ & c > 1. \end{aligned}$$

Here we have used

$$A = \max(0, \tfrac{3}{4}[a - c - 2]).$$

It may be checked that (A1) may be applied to all the 24 terms in the coefficient of $h_2(\mathbf{r})$ in (7).

[1] L. D. Faddeev, *Mathematical Aspects of the Three Body Problem* (Daniel Davey Inc., New York, 1965).
[2] L. V. Kantorovich and G. P. Akilov, *Functional Analysis in Normed Spaces* (Macmillan Co., New York, 1964).

Symmetry mappings of constrained dynamical systems and an associated related integral theorem

Jack Levine* and Gerald H. Katzin

Departments of Mathematics and Physics, North Carolina State University, Raleigh, North Carolina*
(Received 20 June 1973)

By use of Lie derivatives symmetry mappings of constrained conservative dynamical systems are formulated in terms of continuous groups of infinitesimal transformations within the configuration space. Such symmetries are called "natural trajectory collineations" in that the total energy has the same fixed value along each trajectory of the natural family, this value being preserved by the symmetry. It is found that these natural trajectory collineations must be conformal motions subject to an additional restriction dependent upon the potential. The corresponding groups of natural trajectory collineations are obtained for a flat configuration space with potential energy functions with rotational invariance about a point. A specialization of the theory to an indefinite Riemannian space-time shows that homothetic transformations are necessary and sufficient to map a natural family of time (space)-like geodesics into itself. A related integral theorem for constrained dynamical systems admitting linear or quadratic constants of the motion is obtained and illustrated. This theorem shows that in general a new constant of the motion will be obtained by deformation of an existing constant of the motion under a natural trajectory collineation.

1. INTRODUCTION

The related integral theorem[1,2,3] is a unified method which shows that if a dynamical system admits a dynamical symmetry, then in general a new constant of the motion will result from the deformation of a given one under this symmetry mapping. In Refs. 1 and 3, the related integral theorem is based upon dynamical symmetry mappings of unconstrained systems. As originally formulated[1] the theorem provided a unified method of generating quadratic first integrals for dynamical systems with geodesic trajectories based upon deformation of the metrical quadratic integral under projective collineations. In a later paper[3] an extended version of the related integral theorem was developed. This generalization provided a method for generating additional constants of the motion for conservative dynamical systems governed by the equation[4]

$$A^i \equiv \frac{Dv^i}{dt} + g^{ij}V_{,j} = 0, \qquad v^i \equiv \frac{dx^i}{dt}, \tag{1.1}$$

where $V(x)$ is the potential energy and g_{ij} is the metric of the configuration space V_n. Based upon the deformation of the energy integral under the dynamical symmetrics of (1.1) (such symmetries were defined to be trajectory collineations) additional quadratic constants of the motion were obtained in a systematic manner. Application of the theory to the Kepler problem and three dimensional isotropic harmonic oscillator showed that the well-known Runge-Lenz vector constant of the motion and symmetric tensor constant of the motion could be obtained in a simple, direct manner by this unified approach.

In Ref. 2, though only briefly sketched, the related integral theorem concept was applied to a constrained (null geodesic) system. In the present paper we extend the analysis of constrained dynamical mappings and formulate an associated related integral theorem for dynamical systems (1.1) for which the condition

$$B \equiv \tfrac{1}{2}g_{ij}v^iv^j + V - E_0 = 0 \tag{1.2}$$

serves as a constraint in that E_0 is a prescribed (fixed) constant. Such dynamical systems are of particular physical interest for two cases. For the case of a positive definite fundamental form with metric g_{ij} these equations define a "natural family"[5,6] of trajectories with fixed total energy E_0. For the second case with indefinite fundamental form the Eqs. (1.1), (1.2) with $V = 0$ define constrained geodesic trajectories (time-like, spacelike, or null depending upon the value of E_0) in a Riemannian spacetime.

We initially deal with the formulation of the conditions for constrained dynamical mappings (Secs. 2 and 3). For the first case we define "natural trajectory collineations" as those dynamical symmetry mappings which do not violate the constraint (1.2). We obtain necessary and sufficient conditions for natural trajectory collineations and find such mappings must be conformal motions in the configuration space of the problem and in addition must satisfy a further restriction which is dependent upon the potential energy function.[7] We then determine (Sec. 4) in a flat configuration space all potential energy functions with rotational invariance about a point and their corresponding groups of natural trajectory collineations.

In Sec. 5 we consider the second case and continue the analysis of constrained dynamical mappings. It is shown that the most general dynamical symmetry mappings which map a natural family of time- (space)like geodesics into itself are homothetic motions.

In Sec. 6 the conditions for the existence of linear and quadratic constants of the motion of a constrained system (1.1), (1.2) are obtained and a related integral theorem applicable to such constants of the motion is formulated. Linear dependency relations between first integrals derived by this method are shown to be essentially related to the structure of the group of constrained dynamical symmetrical mappings.

We conclude the paper with an application of the related integral theorem to the case of null geodesics in a conformally flat Ricci symmetric space by obtaining a derived quadratic constant of the motion.

2. NATURAL TRAJECTORY COLLINEATIONS

Consider a conservative system whose motion equations in configuration space V_n are given by (1.1), where $V = V(x^i, \ldots, x^n)$ is the potential, and g_{ij} define the V_n metric.

A natural family of trajectories of (1.1) is defined by the energy (first) integral (1.2) where E_0 is a prescribed constant. Such a natural family is said to admit

a natural trajectory collineation defined by a vector field $\xi^i(x)$ and a scalar $\phi(x)$ is the infinitesimal transformation (δa = infinitesimal)

$$\bar{x}^i = x^i + \xi^i(x)\delta a \tag{2.1}$$

along with the associated change in differential path parameter

$$d\bar{t} = \{1 + 2\phi[x(t)]\delta a\}dt \tag{2.2}$$

(where ϕ is evaluated along trajectories) maps the natural family of trajectories into itself in that

$$C \equiv \pounds_\xi B = \pounds_\xi(\tfrac{1}{2}g_{ij}v^iv^j + V - E_0) = 0, \tag{2.3}$$

$$D^i \equiv \pounds_\xi A^i = \pounds_\xi\left(\frac{Dv^i}{dt} + g^{ij}V_{,j}\right) = 0, \tag{2.4}$$

for all values of the v^i for which (1.2) is satisfied. In (2.3) and (2.4) the Lie derivative $\pounds_\xi$ is taken with respect to ξ^i.[8]

Equations (2.3), (2.4) lead to the conditions[9]

$$C = \pounds_\xi B = \tfrac{1}{2}(h_{ij} - 4\phi g_{ij})v^iv^j + V_{,i}\xi^i = 0, \tag{2.5}$$

$$D^i = \pounds_\xi A^i = (\pounds_\xi\left\{{i \atop jk}\right\} - \delta^i_j\phi_{,k} - \delta^i_k\phi_{,j})v^jv^k + 4\phi g^{ij}V_{,j} + g^{ij}V_{;jk}\xi^k - g^{jk}V_{,j}\xi^i_{;k} = 0 \tag{2.6}$$

for all values of the v^i's for which (1.2) is satisfied. In (2.5), (2.6)

$$h_{ij} \equiv \pounds_\xi g_{ij} = \xi_{i;j} + \xi_{j;i}, \qquad 2\phi \equiv \pounds_\xi dt/dt. \tag{2.7}$$

With reference then to (2.5) we have by Hilbert's zero-theorem[10] that $C^p \equiv \Phi B$ for some positive integer p, where $\Phi \equiv \Phi(x^1, \ldots, x^n; v^1, \ldots, v^n)$ is a polynomial of degree $2p-2$ in the v^i. We show next that $p = 1$.

Suppose first $n \geq 3$. Then B considered as a polynomial in the v^i must be irreducible. For if not, and we have $B \equiv (\lambda_i v^i + \lambda_0)(\mu_j v^j + \mu_0)$ (where the λ's and μ's are functions of the x^i), then by (1.2) we must have $g_{ij}v^iv^j \equiv (\lambda_i v^i)(\mu_j v^j)$ which implies the matrix $[g_{ij}]$ is of rank[11] ≤ 2, a contradiction. Since B is irreducible ($n \geq 3$) it follows from the relationship between B and C^p that B is a factor of C, i.e., we may write $C = \rho B$, $[\rho = \rho(x^1, \ldots, x^n)]$, implying $p = 1$.

Suppose next $n = 2$, and B is reducible, $B \equiv LM$, a product of two linear factors (in v^1, v^2), with $L \neq \alpha M$ $[\alpha = \alpha(x^1, x^2)]$. Then $C^p = \Phi B$ implies $C = \rho LM = \rho B$. If $L = \alpha M$, $B = \alpha L^2$, then it can be shown $|g_{ij}| = 0$, a contradiction. If B is irreducible, it follows as in case $n \geq 3$, that $p = 1$. Finally, if $n = 1$, it can be shown, as for $n = 2$, that $C = \rho B$, so $p = 1$.

In a like manner we have $D^i = \psi^i B$, $[\psi^i = \psi^i(x^i, \ldots, x^n)]$ for all n.

It follows from the above-derived relations between C, B and D, B, and (1.2), (2.5), (2.6), (2.7) that

$$\pounds_\xi g_{ij} = 2\mu g_{ij} \quad (2\mu \equiv \rho + 4\phi), \tag{2.8}$$

$$Y \equiv V_{,i}\xi^i - \rho(V - E_0) = 0, \tag{2.9}$$

$$\pounds_\xi\left\{{i \atop jk}\right\} = \delta^i_j\phi_{,k} + \delta^i_k\phi_{,j} + \tfrac{1}{2}\psi^i g_{jk}, \tag{2.10}$$

$$W^i \equiv 4\phi g^{ij}V_{,j} + g^{ij}V_{;jk}\xi^k - g^{jk}V_{,j}\xi^i_{;k} - (V - E_0)\psi^i = 0. \tag{2.11}$$

3. NATURAL TRAJECTORY COLLINEATIONS (CONTINUED)

We consider in further detail the four equations (2.8)-(2.11) as necessary conditions for a natural trajectory collineation. The condition (2.8) implies the mapping vector ξ^i defines a conformal motion, and it is known that for such a symmetry[12]

$$\pounds_\xi\left\{{i \atop jk}\right\} = \delta^i_j\mu_{,k} + \delta^i_k\mu_{,j} - g_{jk}g^{im}\mu_{,m}. \tag{3.1}$$

Equations (2.10) and (3.1) imply

$$\delta^i_j\lambda_{,k} + \delta^i_k\lambda_{,j} - g_{jk}g^{im}\mu_{,m} = \tfrac{1}{2}\psi^i g_{jk} \quad (\lambda = \mu - \phi). \tag{3.2}$$

On putting $i = j$ in (3.2) and summing, we find $\psi_i \equiv g_{ij}\psi^j$ is a gradient,

$$\psi_i \equiv \psi_{,i} = (2n\lambda - 2\phi)_{,i} \qquad (\psi = 2n\lambda - 2\phi + c_0, c_0 = \text{const}). \tag{3.3}$$

If (3.2) be multiplied by $g_{hi}g^{jk}$ and summed on i, j, k there is obtained by use of (3.3), $(n-1)(n+2)\lambda_{,h} = 0$. Hence if $n > 1$, $\lambda = \mu - \phi = c = \text{const}$. We can thus express μ, ρ [see (2.8)], ψ, and ψ^i in terms of ϕ,

$$\mu = \phi + c, \quad \rho = 2(c - \phi), \quad \psi = -2\phi + c', \quad \psi^i = 2\phi_{,j}g^{ij} \qquad (c, c' = \text{const}). \tag{3.4}$$

Hereafter, we take $n > 1$. If then we form the expression $g_{im}W^i$ from (2.11) and make use of (2.7), (2.8) and the relation $g_{ij}\xi^i_{;k} = \xi_{j;k}$, we obtain

$$g_{im}W^i = -\rho V_{,m} + V_{;j}\xi^j_{;m} + V_{;km}\xi^k - (V - E_0)\psi_{,m}. \tag{3.5}$$

Now by the use of (2.7), (2.8), (2.9), (3.4), (3.5) we find $g_{im}W^i = Y_{,i}$. This implies (2.11) is a consequence of (2.8), (2.9), (3.4). In addition, a simple calculation shows that (2.10) is a consequence of (2.8), (3.4). By means of (3.4) we can rewrite (2.8), (2.9) in the form

$$K_{ij} \equiv \xi_{i;j} + \xi_{j;i} - 2(\phi + c)g_{ij} = 0, \tag{3.6}$$

$$Z \equiv V_{,i}\xi^i + 2(\phi - c)(V - E_0) = 0, \tag{3.7}$$

and hence (3.6), (3.7) are necessary conditions for a natural trajectory collineation [see (2.5), (2.6)]. It is easy to show (3.6), (3.7) will imply (2.5), (2.6) [making use of the dependency of (2.10), (2.11) on (3.6), (3.7), and relations (3.4)].

The above results can now be summarized in the theorem to follow:

Theorem 3.1: Equations (3.6) *and* (3.7) ($n > 1$), *are necessary and sufficient conditions for the infinitesimal transformation* (2.1) *with the associated change in differential parameter* (2.2) *to define a natural trajectory collineation of a constrained dynamical system* (1.1), (1.2).

4. NATURAL TRAJECTORY COLLINEATIONS FOR POTENTIALS HAVING ROTATIONAL INVARIANCE

As an application of Theorem 3.1 we determine the nature of the natural trajectory collineations when the configuration space is a Euclidean space $S_n (n \geq 3)$, and $V = V(r)$, where $r^2 = (x^1)^2 + \cdots + (x^n)^2$ (x^i are rectangular coordinates, so $g_{ij} = \delta_{ij}$). It is found convenient to use (2.8), (2.9) [rather than the equivalent (3.6), (3.7)].

As previously observed (2.8) imply that vector ξ^i defines a conformal motion. Now the form for such a ξ^i in an S_n referred to the (x^i) rectangular coordinates is known to be[13,14]

$$\xi^i = b^i + \mu x^i - \tfrac{1}{2}a_i r^2 + b^i_j x^j \quad (b^j_j + b^j_i = 0), \quad (4.1)$$

where

$$\mu = a_0 + \alpha, \quad \alpha \equiv a_i x^i \quad (a_0, a_i, b^i, b^i_j \text{ const.}) \quad (4.2)$$

[the μ is the same as in (2.8)]. It is known that vector ξ^i generates the general conformal group G_N of $N = (n+1)(n+2)/2$ parameters[15]

$$G_N = [U, P_i, V_i, S_{ij}] \quad (i,j = 1, 2, \ldots, n), \quad (4.3)$$

where $(p_i \equiv \partial/\partial x^i)$

$$U = x^i p_i, \quad P_i = p_i, \quad V_i = x^i x^j p_j - \tfrac{1}{2}r^2 p_i, \quad S_{ij} = x^i p_j - x^j p_i. \quad (4.4)$$

There remains the solution of (2.9) for $V = V(r)$, using $\rho = 4c - 2\mu$ [as determined from (3.4)], and using (4.1) for ξ^i.

We first dispose of the two cases $V = E_0$ and $V = V_0 \neq E_0$ ($V_0 = \text{const}$). If $V = E_0$, then (2.9) is satisfied identically and hence the group of natural trajectory collineations is G_N itself. If $V = V_0 \neq E_0$, (2.8) shows $\rho = 4c - 2\mu = 0$, so $\mu = 2c = \text{const}$. This implies $a_i = 0$, and the group of natural trajectory collineations is the subgroup $[U, P_i, S_{ij}]$ of G_N with $1 + n(n+1)/2$ parameters.

Assume now that $V(r) \neq \text{const}$. Define $W \equiv \ln(V - E_0)$. Then (2.9) may be written

$$(W'/r)\sum x^i \xi^i = 4c - 2\mu \quad (W' = dW/dr \neq 0), \quad (4.5)$$

which by (4.1) reduces to

$$X[\beta + (2a_0 + \alpha)r^2/2] + 2\alpha = 4c - 2a_0 \quad (\beta \equiv \sum b^i x^i, \quad X \equiv W'/r \neq 0). \quad (4.6)$$

If (4.6) be differentiated with respect to x^i the result may be written $Px^i + R_i = 0$, where

$$P \equiv X'[\beta + (2a_0 + \alpha)r^2/2]/r + (2a_0 + \alpha)X, \quad R_i \equiv X(b^i + a_i r^2/2) + 2a_i = R_i(r). \quad (4.7)$$

Suppose it were possible that one of the R_i's is 0 and another does not equal 0. Then condition $Px^i + R_i = 0$ will lead to a contradiction. This implies all $R_i = 0$, or all $R_i \neq 0$. If all $R_i \neq 0$ (so $P \neq 0$), then $Px^i + R_i = 0$ gives $x^i/x^j = R_i/R_j$, which again leads to a contradiction. Hence we must have all $R_i = 0$. If we form the equation $XR_{i,j} = 0$, and make use of $R_i = 0$, we obtain

$$(2X' - rX^2)a_i = 0. \quad (4.8)$$

Before discussing (4.8) we consider first the possibility that (4.6) is satisfied identically, i.e., assume

$$\beta + (2a_0 + \alpha)r^2/2 = 0, \quad 4c - 2a_0 - 2\alpha = 0. \quad (4.9)$$

The second of (4.9) implies $\alpha = 0$, $a_0 = 2c$ ($\alpha = 0$ implies $a_i = 0$). Hence the first of (4.9) implies $\beta = 0$, $a_0 = 0$, i.e., we must have $a_0 = a_i = b^i = 0$ ($c = 0$), and the corresponding group of natural trajectory collineations is $[S_{ij}]$, and $V(r)$ is arbitrary.

We return now to (4.8) and assume first $2X' - rX^2 = 0$, which gives $X = -4(C_0 + r^2)^{-1}$, and so $V - E_0 = C_1(C_0 + r^2)^{-2}$ ($C_0, C_1 = \text{const}, C_1 \neq 0$). From (4.6) we thus obtain

$$a_0 C_0 = 0, \quad b^i = \tfrac{1}{2}C_0 a_i, \quad a_0 = -2c. \quad (4.10)$$

In the case $C_0 = 0$, then $V - E_0 = C_1 r^{-4}$, and the group is $[U, V_i, S_{ij}]$ of $1 + n(n+1)/2$ parameters. In the case $C_0 \neq 0$, then $a_0 = 0$ ($c = 0$), and the group is $[\frac{1}{2}C_0 P_i + V_i, S_{jk}]$ of $n(n+1)/2$ parameters.

Finally, we assume from (4.8) that $2X' - rX^2 \neq 0$, so that $a_i = 0$. Then $R_i = 0$ shows $b^i = 0$ (recall $X \neq 0$). Hence (4.6) gives $a_0 W' = (4c - 2a_0)/r$, leading to $V - E_0 = C_1 r^m$ ($m \neq -4$, as then $2X' - rX^2 = 0$). The group of natural trajectory collineations is now $[U, S_{ij}]$ of $1 + n(n-1)/2$ parameters.

The results of this section are summarized in Theorem 4.1 stated below. In this theorem if $[X_1, X_2, \ldots, X_q]$ is any stated group of natural trajectory collineations the notation $\mu(X_\theta)$ means the value of μ corresponding to the generator X_θ. A like meaning applies to $\phi(X_\theta)$ and $c(X_\theta)$ [$\phi(X_\theta) = \mu(X_\theta) - c(X_\theta)$]. In all cases of the theorem the μ-values for the basic generators U, P_i, V_i, S_{ij} are, respectively,

$$\mu(U) = 1, \quad \mu(P_i) = 0, \quad \mu(V_i) = x^i, \quad \mu(S_{ij}) = 0. \quad (4.11)$$

Hence, for example, in the second type $V = V_0$ it was shown above that $\mu = 2c$. It follows that in this case $\phi(U) = 1 - c(U) = 1 - \frac{1}{2}\mu(U) = \frac{1}{2}$.

Theorem 4.1: If the infinitesimal mapping; (2.1), (2.2) *defines a natural trajectory collineation of a conservative dynamical system,* (1.1), (1.2) *(with a Euclidean configuration space* $n \geq 3$*), whose potential is* $V = V(r)$ $[r^2 = \sum (x^i)^2]$*, then* $V(r)$ *and the corresponding group of natural trajectory collineations will be*

TABLE I.[a] Natural trajectory collineation groups for rotationally invariant potentials.

Type	Potential	Group	$\phi(X)$
I	$V = E_0$	G_N	$\phi(U) = c_0$, $\phi(P_i) = c_i$, $\phi(V_i) = x^i - c'_i$, $\phi(S_{ij}) = c_{ij}$, c_0, c_i, c'_i, c_{ij} consts
II	$V = V_0 \neq E_0$, V_0 const	$[U, P_i, S_{ij}]$	$\phi(U) = \frac{1}{2}$, $\phi(P_i) = 0$, $\phi(S_{ij}) = 0$
III	$V(r)$ arbitrary	$[S_{ij}]$	$\phi(S_{ij}) = 0$
IVa	$V - E_0 = C_0 r^{-4}$, C_0 const $\neq 0$	$[U, V_i, S_{ij}]$	$\phi(U) = \frac{3}{2}$, $\phi(V_i) = x^i$, $\phi(S_{ij}) = 0$
IVb	$V - E_0 = C_1(C_0 + r^2)^{-2}$, C_0, C_1 consts $\neq 0$	$[\frac{1}{2}C_0 P_i + V_i, S_{ij}]$	$\phi(\frac{1}{2}C_0 P_i + V_i) = x^i$, $\phi(S_{ij}) = 0$
V	$V - E_0 = C_0 r^m$, $m \neq -4$	$[U, S_{ij}]$	$\phi(U) = c_0$, $\phi(S_{ij}) = 0$

[a]The notation in this table follows that of Section 4.

one of the forms given in Table I. The groups are all subgroups (or the full group) of the general conformal group

5. CONSTRAINED GEODESIC MAPPING IN RELATIVITY THEORY

We now apply the results of the previous sections to the problem of mapping constrained geodesics in relativity theory. We therefore consider the V_n to be a Riemannian space V_4 with signature -2 and take $V = 0$ in (1.1), (1.2), and subsequent equations leading to (3.6), (3.7). [Note that the derivation of (3.6), (3.7) is independent of the signature of the quadratic form in (1.2).] It then follows that the natural trajectory collineation conditions (3.6), (3.7) may be interpreted as the conditions for mapping a natural family of geodesics into itself.

For the specific case in which $V = E_0 = 0$, and t is a suitably chosen parameter, (1.1), (1.2) will define null geodesics. By inspection of (3.6) and (3.7), we obtain the well-known necessary and sufficient conditions[2] that the transformation (2.1), (2.2) maps null geodesics into null geodesics, i.e., the mapping must be a general conformal motion as shown by (3.6).

We next assume that $E_0 \neq 0$ ($V = 0$). Equations (1.1), (1.2) now define timelike ($E_0 > 0$) or spacelike ($E_0 < 0$) geodesics.[16] Equations (3.6) and (3.7) may then be interpreted as necessary and sufficient conditions that (2.1), (2.2) map a natural family of time- (space)like geodesics into itself. Inspection of (3.6) and (3.7) shows such mappings must be homothetic.[17] Hence we may state

Theorem 5.1: A necessary and sufficient condition that the infinitesimal transformation (2.1), (2.2) *maps a natural family of time- (space)like geodesics into itself is that the mapping to be a homothetic motion.*

Since homothetic motions are also conformal, it follows that such mappings are sufficient to take null geodesics into null geodesics.

For the particular case of a Minkowski space-time we note that the homothetic motion group consists of the 10-parameter inhomogeneous Lorentz group plus a scale change.

6. RELATED INTEGRAL THEOREM BASED ON NATURAL TRAJECTORY COLLINEATIONS

In this section we shall develop a related integral theorem[1,2,3] for dynamical systems (1.1) with quadratic constraint (1.2). [As is well-known[1,2,3] this type of theorem will generate additional first integrals (referred to as derived first integrals) from given first integrals.] We consider first the conditions that such systems admit linear or quadratic constants of the motion.

For the existence of a linear constant of the motion it is required that

$$L \equiv A_i(x)v^i = \text{const} \tag{6.1}$$

for all $x^i(t)$ which satisfy (1.1) subject to the constraint (1.2). The requirement $DL/dt = 0$, with use of (1.1), leads to the condition

$$A_{i;j}v^iv^j - A^jV_{,j} = 0, \qquad A^j \equiv g^{ji}A_i, \tag{6.2}$$

subject to the constraint (1.2). By an argument similar to that of Sec. 2 it then follows that necessary and sufficient conditions for a linear first integral of a natural trajectory are

$$E_{ij} \equiv A_{i;j} + A_{j;i} - (2/n)A^m_{;m}g_{ij} = 0, \tag{6.3}$$

$$F \equiv A^jV_{,j} + (2/n)(V - E_0)A^m_{;m} = 0. \tag{6.4}$$

For the existence of a quadratic constant of the motion

$$J \equiv B_{ij}(x)v^iv^j + \Omega(x) = \text{const} \quad (B_{ij} = B_{ji}) \tag{6.5}$$

of a natural trajectory we obtain in a like manner the necessary and sufficient conditions[18]

$$G_{ijk} \equiv B_{(ij;k)} - (1/2)\Gamma_{(k}g_{ij)} = 0, \tag{6.6}$$

$$H_j \equiv -2g^{ik}B_{ij}V_{,k} + \Omega_{,j} - \Gamma_j(V - E_0) = 0, \tag{6.7}$$

where [from (6.6)] it follows that

$$\Gamma_k = 2(n+2)^{-1}(B^i_{i;k} + 2B^i_{k;i}), \qquad B^i_j \equiv g^{ik}B_{kj}. \tag{6.8}$$

We now proceed with the formulation of a related integral theorem for the constrained dynamical system (1.1), (1.2). Assume such a system admits a linear first integral L and there exists a natural trajectory collineation defined by a vector ξ^i as described in Theorem 3.1. By forming $\pounds_\xi L$ with respect to the natural trajectory collineation vector, we obtain

$$\pounds_\xi L = a_iv^i, \qquad a_i \equiv \pounds_\xi A_i - 2\phi A_i. \tag{6.9}$$

We shall now show that [refer to equations (6.3) and (6.4)]

$$e_{ij} \equiv a_{(i;j)} - (2/n)a^m_{;m}g_{ij} = 0, \qquad a^i \equiv g^{ij}a_j, \tag{6.10}$$

$$f \equiv a^jV_{,j} + (2/n)(V - E_0)a^m_{;m} = 0, \tag{6.11}$$

and thus verify that $\pounds_\xi L$ is a constant of the motion. The expansion of (6.10) by use of (6.9) gives

$$\begin{aligned} e_{ij} = {} & (\pounds_\xi A_i)_{;j} + (\pounds_\xi A_j)_{,i} - 2\phi(A_{i;j} + A_{j;i}) \\ & - 2(\phi_{,i}A_j + \phi_{,j}A_i) - (2/n)[g^{ab}(\pounds_\xi A_a)_{;b} \\ & - 2A^a\phi_{,a} - 2\phi A^m_{;m}]g_{ij}. \end{aligned} \tag{6.12}$$

By use of (3.6) we evaluate $\pounds_\xi\{^i_{jk}\}$ and substitute in the identity[12]

$$(\pounds_\xi A_i)_{;j} \equiv \pounds_\xi(A_{i;j}) + (\pounds_\xi\{^m_{ij}\})A_m \tag{6.13}$$

to obtain

$$(\pounds_\xi A_i)_{;j} = \pounds_\xi(A_{i;j}) + \phi_{,j}A_i + \phi_{,i}A_j - g_{ij}A^m\phi_{,m}. \tag{6.14}$$

Use of (6.14), (3.6), and (6.3) in (6.12) leads to $e_{ij} = 0$.

To verify $f = 0$ we expand (6.11) by use of (6.9) and (6.14) to obtain

$$\begin{aligned} f = {} & g^{ij}(\pounds_\xi A_i)V_{,j} - 2\phi A^iV_{,i} + (2/n)(V - E_0)g^{ab}\pounds_\xi(A_{a;b}) \\ & - 2(V - E_0)A^a\phi_{,b} - (4/n)(V - E_0)A^m_{;m}. \end{aligned} \tag{6.15}$$

We observe that f as given by (6.15) may be expressed in the form

$$\begin{aligned} f = {} & \pounds_\xi F - (2/n)A^m_{;m}Z - A^mZ_{;m} + K_{ij}A^ig^{kj}V_{,k} + 2\phi F \\ & + g^{ij}[(V - E_0)/n][\pounds_\xi E_{ij} + (2/n)E_{ij}] = 0, \end{aligned} \tag{6.16}$$

where the symbols E_{ij}, F, K_{ij}, and Z are defined by (6.3), (6.4), (3.6), and (3.7), respectively. This last remark completes the verification that $\pounds_\xi L$ is a constant of the motion.

We next consider the verification that $\pounds_\xi J$ is a constant of the motion, where $\pounds_\xi J$ is based upon the deformation under a natural trajectory collineation of a given quadratic first integral J of the form (6.5). We form

$$\pounds_\xi J \equiv b_{ij}v^iv^j + \theta, \qquad b_{ij} \equiv \pounds_\xi B_{ij} - 4\phi B_{ij}, \qquad \theta \equiv \pounds_\xi \Omega, \tag{6.17}$$

and define

$$h_{ijk} \equiv b_{(ij;k)} - \tfrac{1}{2}\chi_{(k}g_{ij)}, \tag{6.18}$$

$$m_j \equiv -2b_{ij}g^{ki}V_{,k} + \theta_{,j} - \chi_j(V - E_0), \tag{6.19}$$

where

$$\chi_k \equiv 2(n+2)^{-1}(b^i_{i;k} + 2b^i_{k;i}), \qquad b^i_k \equiv g^{ij}b_{jk}. \tag{6.20}$$

We shall now show [refer to (6.6), (6.7), and (6.8)] that h_{ijk} and m_j vanish identically for b_{ij} and θ defined by (6.17) provided (6.6), (6.7), (6.8), (3.6), and (3.7) are satisfied.

If (3.6) is used in an identity similar to (6.3), we obtain

$$(\pounds_\xi B_{ab})_{;k} = \pounds_\xi(B_{ab;k}) + 2B_{ab}\phi_{,k} + B_{kb}\phi_{,a} + B_{ka}\phi_{,b} - (g_{ak}B^m_b + g_{bk}B^m_a)\phi_{,m}. \tag{6.21}$$

We expand χ_k of (6.20) by use of (6.17) and (6.21) and make use of (6.6), (6.8), and (3.6) to obtain

$$\chi_k = \pounds_\xi \Gamma_k - 2\phi\Gamma_k - 4\phi_{,m}B^m_k + 2c\Gamma_k. \tag{6.22}$$

If the right side of (6.18) is expanded by use of (6.17) and (6.22) and use be made of (6.21), (3.6), and (3.7) we find $h_{ijk} = 0$.

Next we show that m_j of (6.19) vanishes. By use of (6.17) and (6.22) we expand (6.19) to obtain

$$m_j = -2(\pounds_\xi B_{ij})g^{ik}V_{,k} + 8\phi B^i_j V_{,i} + (\Omega_{,i}\xi^i)_{,j} - (\pounds_\xi \Gamma_j - 2\phi\Gamma_j - 4B^i_j\phi_{,i} + 2c\Gamma_j)(V - E_0). \tag{6.23}$$

From (6.7) we form $Y_j \equiv \pounds_\xi H_j$ and made use of (3.6) to obtain

$$Y_j = -2(\pounds_\xi B_{ij})g^{ik}V_{,k} + 4(\phi + c)B^k_j V_{,k} - 2B^k_j(V_{,m}\xi^m)_{,k} + (\Omega_{,i}\xi^i)_{,j} - (\pounds_\xi \Gamma_j)(V - E_0) - \Gamma_j V_{,i}\xi^i = 0. \tag{6.24}$$

By use of (3.7) and (6.24) we may express (6.23) in the form

$$m_j = Y_j + 2Z_{,k}B^k_j + \Gamma_j Z = 0. \tag{6.25}$$

This completes the verification that $\pounds_\xi J$ defined by (6.17) is a first integral of the constrained dynamical system.

We may summarize by stating the following Related Integral Theorem of a constrained dynamical system.

Theorem 6.1: If a family of natural trajectories [defined to be solutions of the dynamical equation (1.1) subject to the quadratic constraint (1.2)] admits a quadratic first integral (6.5) [linear first integral (6.1)], and if there exists a natural trajectory collineation based upon vector ξ^i as described by Theorem 3.1, then in general the family of natural trajectories will admit an additional quadratic first integral (6.17) [linear first integral (6.9)], which is based upon the deformation of the original quadratic (linear) integral under the trajectory collineation.

If in Theorem 6.1 we take the V_n to be of indefinite signature and choose $V = 0$, $E_0 = 0$ [in (1.1) and (1.2)], then Theorem 6.1 may be regarded as a related integral theorem for null geodesics. If, in addition, we take $\Omega = 0$ [in (6.5)], we then obtain Theorem 4.1 of Ref. 2.

Next we obtain dependency relations between derived first integrals. We assume the existence of an r-parameter group of natural trajectory collineations defined by the vectors ξ^i_α, $\alpha = 1, \ldots, r$. Based upon a given linear or quadratic first integral I and a natural trajectory vector ξ^i_α, we obtain a "first derived" first integral $I_\alpha \equiv \pounds_\alpha I$ ($\pounds_\alpha \equiv \pounds_{\xi_\alpha}$) by use of Theorem 6.1. Repeated application of Theorem 6.1 based upon vector ξ^i_β in general leads to a "second derived" first integral $I_{\beta\alpha} \equiv \pounds_\beta \pounds_\alpha I$.

We now assume $\alpha \neq \beta$. By use of the group relationship[12]

$$\pounds_\alpha \pounds_\beta - \pounds_\beta \pounds_\alpha = C^\gamma_{\alpha\beta}\pounds_\gamma, \tag{6.26}$$

where $C^\gamma_{\alpha\beta}$ are the structure constants of the natural trajectory collineation group, we immediately obtain the linear dependency relations

$$I_{\beta\alpha} - I_{\alpha\beta} = C^\gamma_{\beta\alpha}I_\gamma \tag{6.27}$$

between first integrals.

From (3.7) based upon vector ξ^i_α and associated constant c_α we easily find

$$C^\gamma_{\alpha\beta}c_\gamma = 0. \tag{6.28}$$

Use of (6.28) and the group relationship[12]

$$\pounds_\beta \xi^i_\alpha = C^\gamma_{\beta\alpha}\xi^i_\gamma$$

permits a detailed verification of (6.27) based upon given linear or quadratic first integrals I of the form (6.1) or (6.5) (refer to Ref. 3 for details of a similar calculation).

7. ILLUSTRATION OF THE RELATED INTEGRAL THEOREM

In this section we give an illustration of Theorem 6.1 as applied to null geodesics of a V_4 of signature + 2. If in (1.1), (1.2) we take $V = E_0 = 0$, then the family of natural trajectories becomes the null geodesics of the space.

We take the V_4 to be a conformally flat space C_4 defined by the fundamental quadratic form

$$\Phi \equiv \frac{1}{u^2}\sum_1^4 e_i(dx^i)^2, \qquad (e_1, e_2, e_3, e_4) \equiv (1, 1, 1, -1), \tag{7.1}$$

where

$$u^2 \equiv S^2 + Q + \tfrac{1}{4}, \qquad S \equiv \sum_1^4 e_i(x^i)^2, \qquad Q \equiv (x^2)^2 + 2x^1x^3 + (x^4)^2. \tag{7.2}$$

It can be verified for this C_4 that the scalar curvature $R = 0$ and the Ricci tensor is covariant constant, i.e., $R_{ij;k} = 0$.[19]

It follows that if in (6.5) we choose $\Omega = 0$ and $B_{ij} = R_{ij}$, then $R_{ij}v^iv^j$ is a quadratic first integral of the null geodesics since the conditions (6.6), (6.7) are satisfied.[20]

In this C_4 the vector $\xi^i \equiv \delta^i_i$ is a conformal motion vector ($\pounds_\xi g_{ij} = 2\mu g_{ij}$, $4\mu = \xi^i_{;i}$). Based upon this vector we define by (6.17) [with $\phi = \mu - c$ by (3.6)]

$$b_{ij} \equiv (\pounds_\xi R_{ij} - 4\mu R_{ij}) + 4cR_{ij}. \tag{7.3}$$

By Theorem 6.1 as applied to null geodesics, $b_{ij}v^iv^j$ defines a derived first integral of these geodesics.

It is of interest to determine if the quadratic first integral defined by (7.3) is a linear sum of known quadratic first integrals of the null geodesics in this C_4. Two such known quadratic integrals are $g_{ij}v^iv^j$ and $R_{ij}v^iv^j$. In addition, it is known the C_4 admits the full group G_{15} of conformal motions[13,14,15,21] and each conformal motion vector defines a linear first integral of the null geodesics [see (6.3), (6.4)]. The product of any two such linear first integrals defines a quadratic integral of the null geodesics. If the quadratic first integral defined by b_{ij} is in fact a linear sum of the above mentioned quadratic first integrals, then[22]

$$b_{ij} = A_0R_{ij} + B_0g_{ij} + \sum_{\alpha,\beta=1}^{15} K^{\alpha\beta}\xi_i(X_\alpha)\xi_j(X_\beta) \tag{7.4}$$

must hold for some constants (not all zero) $A_0, B_0, K^{\alpha\beta}$, where G^{15} has generators $[X_1, \ldots, X_{15}] = [U, P_i, V_i, S_{ij}]$ [see (4.3)], and $\xi^i(X_\alpha)$ indicates the vector corresponding to generator X_α. It can be shown that (7.4) (using for example, $i = j = 2$) cannot be satisfied for any such choice of these constants, and hence $b_{ij}v^iv^j$ is linearly independent of these known first integrals.

[1] G. H. Katzin and J. Levine, J. Math. Phys. **9**, 8 (1968).
[2] G. H. Katzin and J. Levine, Colloquium Mathematicum (Wroclaw, Poland) **26**, 21 (1972).
[3] G. H. Katzin, J. Math. Phys. **14**, 1213 (1973).
[4] Latin indices run from 1 to n. The Einstein summation notation is used unless stated otherwise. We indicate absolute differentiation with respect to path parameter t by D/dt, total differentiation by d/dt, partial differentiation with x^k by comma (,), and covariant differentiation [based upon Christoffel symbols $\{^i_{jk}\}$] by semicolon (;).
[5] E. Kasner, *Differential – Geometric Aspects of Dynamics, Colloquium Publication* (Amer. Math. Soc., New York, 1913), Vol. III.
[6] E. T. Whittaker, *A Treatise on the Analytical Dynamics of Particles and Rigid Bodies* (Cambridge U. P., Cambridge, 1936).
[7] See Ref. 5 for a related discussion of conformal mappings of natural dynamical systems.
[8] For a general discussion of Lie derivatives see K. Yano, *The Theory of Lie Derivatives and Its Applications* (North-Holland, Amsterdam, 1957).
[9] See Ref.3 for details of similar calculations.
[10] W. Hodge and D. Pedoe, *Methods of Algebraic Geometry* (Cambridge) (U. P., 1968).
[11] M. Bôcher, *Introduction to Higher Algebra* (MacMillan, New York, 1929).
[12] See Ref. 8.
[13] J. Levine and G. H. Katzin, Tensor, New Series, **21**, 319 (1970).
[14] J. Levine, Bull. Amer. Math. Soc. **42**, 418 (1936).
[15] J. Levine, Bull. Amer. Math. Soc. **45**, 766 (1939).
[16] J. L. Anderson, *Principles of Relativity Physics* (Academic, New York, 1967).
[17] We presume this result to be generally accepted, but we are unaware of the existence of a formal mathematical proof of its validity.
[18] The symbol ( ) denotes symmetrization, so that $B_{(ij;k)} \equiv {}^1/_3(B_{ij;k} + B_{jk;i} + B_{ki;j})$ (making use of the symmetry of B_{ij}).
[19] J. Levine and G. H. Katzin, Tensor, New Series, **19**, 317 (1968).
[20] It should be noted that $R_{ij;k} = 0$ implies $R_{ij}v^iv^j$ is a first integral of all geodesics. However, the application being considered here is to null geodesics.
[21] A. H. Taub, Bull. Amer. Math. Soc. **55**, 85 (1948).
[22] Note that the constant $4c$ in (7.3) has been absorbed in the A_0.

Gravitation and the new improved energy–momentum tensor

Hans C. Ohanian

Rensselaer Polytechnic Institute, Troy, New York 12181
(Received 7 March 1972)

We give precise definitions of the weak and strong principles of equivalence and show that the new gravitational theory based on the improved energy-momentum tensor of Callan, Coleman, and Jackiw [Ann. Phys. (N.Y.) 59, 42 (1970)] satisfies both of these principles. As a consequence of the equality between the 4-momentum given by the canonical energy-momentum tensor and the "momentum" given by the pseudotensor that is the source of gravitation, the weak principle is also shown to hold in more general Einstein theories. Investigation of the interactions of a scalar field in the new gravitational theory shows that, besides the familiar long range interaction, there exists a new short range gravitational interaction between any scalar field and other matter.

I. INTRODUCTION

In Lagrangian field theory the canonical energy–momentum tensor is of great importance since the energy and momentum of a system are obtained by integrating components of this tensor over the volume of the system. If we designate the field variables appearing in the Lagrangian $\mathcal{L}$ by $Q_i(x)$, then the canonical tensor is[1]

$$T'_{\mu}{}^{\nu} = Q_{i,\mu}\,\frac{\partial \mathcal{L}}{\partial Q_{i,\nu}} - \delta_{\mu}{}^{\nu}\mathcal{L}, \tag{1}$$

where the repetition of the index i indicates summation over all the field variables. This tensor is conserved:

$$T'_{\mu}{}^{\nu}{}_{,\nu} = 0. \tag{2}$$

The energy–momentum 4-vector is given by

$$P_{\mu} = \int T'_{\mu}{}^{0}\, d^3x. \tag{3}$$

However, $T'_{\mu}{}^{\nu}$ is not the only acceptable energy–momentum tensor. It is always possible to add to $T'_{\mu}{}^{\nu}$ an extra tensor of the form $W_{\mu}^{[\nu\alpha]}{}_{,\alpha}$ to get a different tensor

$$T'_{\mu}{}^{\nu} + W_{\mu}^{[\nu\alpha]}{}_{,\alpha}. \tag{4}$$

This leaves Eqs. (2) and (3) unchanged. The existence of many alternative energy–momentum tensors has one good consequence: We can construct a symmetric canonical tensor by adding some extra terms to the right-hand side of Eq. (1).[2] But the lack of uniqueness of the energy–momentum tensor means that we cannot unambiguously identify the density of energy and momentum. The theory does not determine how the energy and momentum are distrubuted over the volume of the system.

The ambiguity in the energy–momentum tensor is removed if instead of taking as energy–momentum tensor "what is conserved," we take "what is the source of gravitation." Different densities of energy and momentum will give different gravitational fields and, therefore, in any given gravitational theory, these densities must be unambiguous. In any general Einstein theory[3] we define the energy–momentum tensor by

$$-R_{\mu}{}^{\nu} + \tfrac{1}{2}\delta_{\mu}{}^{\nu}R = 8\pi G T_{\mu}{}^{\nu}. \tag{5}$$

This uniquely determines $T^{\mu\nu}$ as

$$T^{\mu\nu} = -\frac{2}{\sqrt{-g}}\,\frac{\delta \mathcal{L}_m}{\delta g_{\mu\nu}} \tag{6}$$

with

$$\frac{\delta}{\delta g_{\mu\nu}} = \frac{\partial}{\partial g_{\mu\nu}} - \frac{\partial}{\partial x^{\alpha}}\,\frac{\partial}{\partial g_{\mu\nu,\alpha}}, \tag{7}$$

where $\mathcal{L}_m$ is that part of the Lagrangian which contains matter variables. We can take the limit $g_{\mu\nu} \to \eta_{\mu\nu}$ to obtain the corresponding tensor of special relativity. If Eq. (6) is used, it is necessary to show that the resulting tensor gives the correct energy–momentum 4-vector, i.e., it must be shown that the volume integrals of $T_{\mu}{}^{0}$ and $T'_{\mu}{}^{0}$ are the same. We will show that this is so in Sec. III. Equation (6) has the further advantage that it directly gives a symmetric tensor.

Although this settles the question of which is the right energy–momentum tensor in principle, it does not in practice because we do not know how the gravitational field variables enter $\mathcal{L}_m$. One often postulates a *minimal coupling principle*: $\mathcal{L}_m$ in the presence of gravitation is obtained from the corresponding $\mathcal{L}_m$ of special relativity (assumed known) by replacing $\eta_{\mu\nu}$ by $g_{\mu\nu}$ and ordinary derivatives by covariant derivatives.[4] We will call the resulting theory the *minimal Einstein theory*. However, general Einstein theories, involving a direct coupling of the curvature tensor with the matter field variables, are not excluded by the experimental evidence.

Callan, Coleman, and Jackiw[2] have proposed that the correct choice for the energy–momentum tensor (in the flat space limit) is the tensor obtained from the (symmetric) canonical tensor $T'_{\mu}{}^{\nu}$ by adding a term $-\frac{1}{6}(\partial_{\mu}\partial^{\nu} - \delta_{\mu}{}^{\nu}\Box^2)\phi^2$ for *each* scalar field ϕ that is present. If we suppose there is only one such scalar field, then the proposed tensor is

$$\Theta_{\mu}{}^{\nu} = T'_{\mu}{}^{\nu} - \tfrac{1}{6}(\partial_{\mu}\partial^{\nu} - \delta_{\mu}{}^{\nu}\Box^2)\phi^2. \tag{8}$$

This tensor, called the new improved tensor, results from Eq. (6) if one adds to the usual minimal coupling Lagrangian an extra term $-\frac{1}{12}\sqrt{-g}R\phi^2$ representing a direct interaction of the scalar field with the curvature invariant R.

This new tensor is an improvement over the old canonical tensor because: (i) In any renormalizable theory the new tensor has finite matrix elements in every order of perturbation theory (matrix elements of the canonical tensor diverge) and (ii) the currents associated with scale and conformal transformations have simpler expressions in terms of the new tensor.

The gravitational interaction of the new tensor has been discussed by Callan, Coleman, and Jackiw.

Their discussion was based on a particular example of a scalar matter field interacting with itself; all other forms of matter were assumed absent. The following is a more general discussion of these interactions and of the role played by the equivalence principles in the new theory. We begin with some definitions of what will be meant by the weak and strong principles of equivalence. We will prove that in any general Einstein theory (including the theory based on the new tensor) the weak principle is satisfied. Essentially this amounts to a proof that the canonical ($T'_\mu{}^\nu$) and gravitational ($T_\mu{}^\nu$) tensors give the same 4-momentum. The examination of the gravitational interactions of a scalar field with other nonscalar matter fields shows that this gravitational interaction possesses a short range component. Finally, we consider the implications that this has for the strong principle of equivalence.

Our discussion is based on the use of c-number fields. Although it is often obvious what the q-number version of many of the following equations should be, the precise definition of gauge invariant products of field operators is a very difficult problem in the quantum theory of gravitation (see, e.g., Ref. 5). We will only touch upon quantum theory in the simple example of the Appendix.

II. THE EQUIVALENCE PRINCIPLES

A distinction has been made by Dicke[6] between the weak principle of equivalence (WPE) and the strong principle of equivalence (SPE). The WPE states that *in a given gravitational field all test particles fall with the same acceleration.* The SPE states that *in all freely falling, nonrotating laboratories the results of any local experiments are the same, independent of the gravitational field surrounding the laboratory.* The first principle asserts that inertial and gravitational effects are locally indistinguishable as far as the free motion of test particles is concerned. The second asserts that these effects are indistinguishable by any experiment. The trouble with the above statement of the WPE is that it does not make clear what is meant by a "test particle." Roughly, a test particle is a "sufficiently small" body. Just what limit must be set on the size depends on the gradients present in the gravitational field which the test particle is supposed *not to feel.* (One would probably also have to require that the gravitational self-energy be small and that the body have no spin.) Similarly, "local" in the statement of the SPE means that the experiment is confined to a "sufficiently small" region.

We will give two precise statements which roughly correspond to the above. We begin by defining the inertial and gravitational mass of an arbitrary system of finite size. The inertial mass M_I is the energy in the rest frame of the system. To define the gravitational mass M_G we take a standard test body of mass $M_{I,S}$ and define its gravitational mass as $M_{G,S} = M_{I,S}$. The gravitational mass M_G of any arbitrary system can then be defined by the acceleration a that the standard test body experiences when released at a large distance r from the system in otherwise empty space:

$$M_G = \lim_{r\to\infty} (ar^2/G). \tag{9}$$

In the particular case where the test body is falling towards an identical copy of itself, Eq. (9) serves to define G:

$$G = \lim_{r\to\infty} (ar^2/M_{I,S}). \tag{10}$$

As a precise statement of the WPE we now take the following: *The gravitational mass of any system equals its inertial mass.*

It is clear that the main difference between this statement of the WPE and the earlier one is that the limiting procedure needed to eliminate tidal forces has been made explicit. Furthermore, no restriction is imposed on the strength of the internal gravitational fields of the system.

As a precise statement of the SPE we take: *At each point of space-time it is possible to find a coordinate transformation such that the gravitational field variables disappear from the field equation of matter.*

This statement is related to the one given at the beginning of this section; but the transformation that eliminates the gravitational fields from the field equations is only formally a coordinate transformation. The transformation will depend on the microscopic gravitational fields, and it therefore cannot be interpreted physically as a transformation between macroscopic reference frames.

It is clear that the minimal coupling principle implies the SPE, since only $g_{\mu\nu}$ and $g_{\mu\nu,\alpha}$ appear in the equations of motion of matter in the minimal Einstein theory. These gravitational variables can be eliminated by a transformation to the local geodesic coordinates.

III. THE WEAK EQUIVALENCE PRINCIPLE IN EINSTEIN THEORIES

We will show that the equality of inertial and gravitational mass holds in all general Einstein theories.

Suppose that the coupling of matter to gravitation is *not* minimal. This means that besides the usual minimal coupling Lagrangian density $\mathfrak{L}_{\min}$, there is an extra term dependent on the curvature tensor

$$\mathfrak{L} = (1/16\pi G)\,\text{“}\sqrt{-g}R\text{”} + \mathfrak{L}_{\min} + \text{“}\sqrt{-g}R_{\alpha\beta\sigma\rho}f^{\alpha\beta\sigma\rho}\text{”}. \tag{11}$$

The expression “$\sqrt{-g}R$” stands for the integrand that results when one integrates $\int\sqrt{-g}Rd^4x$ by parts wherever second derivatives of $g_{\mu\nu}$ appear and omits the surface terms; the definition of “$\sqrt{-g}R_{\alpha\beta\sigma\rho}f^{\alpha\beta\sigma\rho}$” is similar. The function $f^{\alpha\beta\sigma\rho}(q_i)$ is an arbitrary function of the matter variables q_i. These matter variables can include any number of scalar fields, spinor fields, and vector fields, etc. We assume the Lagrangian yields field equations of second differential order: consequently $f^{\alpha\beta\sigma\rho}$ *cannot* depend on the derivatives of the q_i. Examples of nonminimal coupling terms are $\sqrt{-g}R\phi^2$ (where ϕ is a scalar field) and $\sqrt{-g}R_{\mu\nu}\bar{\psi}\sigma^{\mu\nu}\psi$ (where ψ is a spinor field).

The field equation can be written in the form

$$\frac{1}{16\pi G}\frac{\partial}{\partial x^\beta}\left(\frac{g_{\mu\tau}}{\sqrt{-g}}\frac{\partial}{\partial x^\alpha}\left(\mathfrak{g}^{\nu\tau}\mathfrak{g}^{\alpha\beta} - \mathfrak{g}^{\nu\alpha}\mathfrak{g}^{\tau\beta}\right)\right) = \mathfrak{T}_\mu{}^\nu \tag{12}$$

with $\mathfrak{g}^{\mu\nu} = g^{\mu\nu}\sqrt{-g}$. The object $\mathfrak{T}_\mu{}^\nu$, which we call the *gravitational pseudotensor*, is given by

$$\mathfrak{T}_\mu{}^\nu = \mathfrak{t}_\mu{}^\nu + \tau_\mu{}^\nu, \tag{13}$$

where $\mathfrak{t}_\mu{}^\nu$ is the Einstein pseudotensor for the gravi-

tational field, and $\tau_\mu{}^\nu$ is the energy–momentum tensor density of matter

$$\tau^{\mu\nu} = \sqrt{-g}\,T^{\mu\nu} = -2\frac{\delta \mathcal{L}_m}{\delta g_{\mu\nu}}, \tag{14}$$

where

$$\mathcal{L}_m = \mathcal{L}_{\min} + \text{“}\sqrt{-g}R_{\alpha\beta\sigma\rho}f^{\alpha\beta\sigma\rho}\text{”}. \tag{15}$$

It is possible to define other pseudotensors by adding a term $W_\mu{}^{[\nu\alpha]}{}_{,\alpha}$ to *both* sides of Eq. (12). This will not affect the arguments that follow in any essential way. The lack of uniqueness of the pseudotensor means that although the distribution of the energy and momentum of matter is unambiguous, the distribution of energy and momentum of the gravitational fields is not.

We can show that the gravitational mass of the system is related to the gravitational pseudotensor by

$$M_G = \int \mathfrak{T}_0{}^0 d^3x, \tag{16}$$

where the integral is to be taken over all space. To see this, we observe that $\mathfrak{T}_0{}^0$ is a three-divergence (the time derivatives cancel):

$$\mathfrak{T}_0{}^0 = \frac{1}{16\pi G}\frac{\partial}{\partial x^k}\left(\frac{g_{0\tau}}{\sqrt{-g}}\frac{\partial}{\partial x^\alpha}\left(g^{0\tau}g^{\alpha k} - g^{0\alpha}g^{\tau k}\right)\right) \tag{17}$$

and hence

$$\int \mathfrak{T}_0{}^0 d^3x = \frac{1}{16\pi G}\int \frac{g_{0\tau}}{\sqrt{-g}}\frac{\partial}{\partial x^\alpha}\left(g^{0\tau}g^{\alpha k} - g^{0\alpha}g^{\tau k}\right)dS_k, \tag{18}$$

where the surface integral is over a surface at infinity. We now suppose that we are in the rest frame of the system and that no gravitational radiation is being emitted. In that case the solutions of the empty space field equations that apply at large distances from the matter distribution have the asymptotic behavior

$$g_{00} \to 1 - B/r, \qquad g_{kl} \to -\delta_{kl}(1 + B/r), \tag{19}$$

where B is a constant. The right-hand side of Eq. (18) can be evaluated by means of these expressions with the result

$$\int \mathfrak{T}_0{}^0 d^3x = B/2G. \tag{20}$$

To identify the constant B, we observe that in the limit $r \to \infty$, the standard test body will move along the geodesics of $g_{\mu\nu}$.[7] This means that the test body will move as though subjected to a potential $\frac{1}{2}(g_{00} - 1)$. Identifying this potential with the Newtonian expression $-GM_G/r$, as required by our definition of M_G, we obtain

$$B = 2GM_G \tag{21}$$

and hence (16) follows.

The inertial mass, or energy, is given in terms of the canonical energy–momentum tensor $T'_\mu{}^\nu$

$$M_I = \int T'_0{}^0 d^3x = \int\left(Q_{i,0}\frac{\partial \mathcal{L}}{\partial Q_{i,0}} - \mathcal{L}\right)d^3x, \tag{22}$$

where Q_i represents the combined sets of variables q_i and $g_{\mu\nu}$. The difference between M_I and M_G is, therefore, the following:

$$M_I - M_G = \int\left[\left(Q_{i,0}\frac{\partial \mathcal{L}}{\partial Q_{i,0}} - \mathcal{L}\right) - \mathrm{t}_0{}^0 - \tau_0{}^0\right]d^3x \tag{23}$$

$$= \int\left[\left(g_{\mu\nu,0}\frac{\partial}{\partial g_{\mu\nu,0}} - 1\right)\mathcal{L}_g - \mathrm{t}_0{}^0\right]d^3x + \int\left[\left(Q_{i,0}\frac{\partial}{\partial Q_{i,0}} - 1\right)\mathcal{L}_m - \tau_0{}^0\right]d^3x, \tag{24}$$

where $\mathcal{L}_g$ stands for the first term on the right-hand side of Eq. (11).

The integrand of the first integral in (24) is zero because the Einstein pseudotensor $\mathrm{t}_0{}^0$ is precisely the canonical tensor associated with $\mathcal{L}_g$.

The second integral vanishes because of the following equality:

$$\int\left(Q_{i,\mu}\frac{\partial \mathcal{L}_m}{\partial Q_{i,0}} - \delta_\mu{}^0\mathcal{L}_m\right)d^3x = \int \tau_\mu{}^0 d^3x. \tag{25}$$

A general proof of this equality is given in Appendix A. We therefore conclude

$$M_I = M_G \tag{26}$$

in all general Einstein theories.

Equation (25) asserts that the tensor $T^{\mu\nu} = \tau^{\mu\nu}/\sqrt{-g}$ of Eq. (6) satisfies the requirement that it gives the correct (i.e., canonical) 4-momentum.

If we add the integral of the Einstein pseudotensor to each side of Eq. (25), we obtain the simple result

$$\int T'_\mu{}^0 d^3x = \int \mathfrak{T}_\mu{}^0 d^3x. \tag{27}$$

The arguments given in Appendix A are valid not only for general Einstein theories, but for any theory containing a tensor field $g_{\mu\nu}$. For example, the result (27) holds in the scalar tensor theory of Brans–Dicke.[8] But the WPE fails in this theory since Eq. (16) breaks down in the presence of the long range scalar field.[9]

IV. THE SHORT RANGE GRAVITATIONAL INTERACTION

We will now investigate the equation of motion of a scalar field interacting with gravitation and other fields. For simplicity we assume that there is only *one scalar field;* but there can be any number of nonscalar matter fields. The equation of motion of the scalar field ϕ of mass μ is determined by the Lagrangian

$$\mathcal{L}_m = \tfrac{1}{2}\sqrt{-g}\,(g^{\mu\nu}\phi_{,\mu}\phi_{,\nu} - \mu^2\phi^2) - \tfrac{1}{12}\sqrt{-g}\,\phi^2 R + \mathcal{L}_I + \mathcal{L}'_m, \tag{28}$$

where $\mathcal{L}_I$ contains the interactions of the scalar field with all the matter fields (including self-interactions) and $\mathcal{L}'_m$ is the Lagrangian of the other matter fields.

The equation of motion is

$$\phi^{,\mu}{}_{;\mu} + \mu^2\phi + \tfrac{1}{6}\phi R - \frac{1}{\sqrt{-g}}\frac{\partial \mathcal{L}_I}{\partial \phi} = 0. \tag{29}$$

The gravitational field equation implies that

$$R = 8\pi G T \tag{30}$$

and hence

$$\phi^{,\mu}{}_{;\mu} + \mu^2\phi + \frac{4\pi G}{3}\phi T - \frac{1}{\sqrt{-g}}\frac{\partial \mathcal{L}_I}{\partial \phi} = 0. \tag{31}$$

The explicit form of $T^{\mu\nu}$ is

$$T^{\mu\nu} = \frac{-2}{\sqrt{-g}}\frac{\delta \mathcal{L}_m}{\delta g_{\mu\nu}} = \frac{-2}{\sqrt{-g}}\frac{\delta}{\delta g_{\mu\nu}}(\mathcal{L}_I + \mathcal{L}'_m) + \phi^{,\mu}\phi^{,\nu} - \tfrac{1}{2}g^{\mu\nu}(\phi^{,\alpha}\phi_{,\alpha} - \mu^2\phi^2) - \tfrac{1}{6}(\phi^{2,\mu;\nu} - g^{\mu\nu}\phi^{2,\alpha}{}_{;\alpha}) - \tfrac{1}{6}\phi^2(R^{\mu\nu} - \tfrac{1}{2}g^{\mu\nu}R), \tag{32}$$

so that

$$T = \frac{-2}{\sqrt{-g}} g_{\mu\nu} \frac{\delta}{\delta g_{\mu\nu}} (\mathcal{L}_I + \mathcal{L}_m') + 2\mu^2\phi^2 + \phi\phi_{;\alpha}^{\cdot\alpha} + \tfrac{1}{6}\phi^2 R = \frac{-2}{\sqrt{-g}} g_{\mu\nu} \frac{\delta}{\delta g_{\mu\nu}} (\mathcal{L}_I + \mathcal{L}_m') + \mu^2\phi^2 + \phi \frac{1}{\sqrt{-g}} \frac{\partial \mathcal{L}_I}{\partial \phi}. \quad (33)$$

Inserting this in Eq. (31) we obtain

$$\phi^{\cdot\mu}{}_{;\mu} + \mu^2\phi - \frac{1}{\sqrt{-g}} \frac{\partial \mathcal{L}_I}{\partial \phi} + \frac{4\pi G}{3\sqrt{-g}} \times \phi\left(\sqrt{-g}\,\mu^2\phi^2 + \phi \frac{\partial \mathcal{L}_I}{\partial \phi} - 2g_{\mu\nu} \frac{\delta}{\delta g_{\mu\nu}} (\mathcal{L}_I + \mathcal{L}_m')\right) = 0. \quad (34)$$

As pointed out in Ref. 2, the term $\sim G\mu^2\phi^3$ is of little interest. It represents a self-interaction which can be eliminated by introducing a counterterm in $\mathcal{L}_I$. We can write

$$\mathcal{L}_I = \mathcal{L}_I' + (\pi G/3)\sqrt{-g}\,\mu^2\phi^4 \quad (35)$$

and then

$$\phi^{\cdot\mu}{}_{;\mu} + \mu^2\phi - \frac{1}{\sqrt{-g}} \frac{\partial \mathcal{L}_I'}{\partial \phi} + \frac{4\pi G}{3\sqrt{-g}} \times \phi\left(\phi \frac{\partial \mathcal{L}_I'}{\partial \phi} - 2g_{\mu\nu} \frac{\delta}{\delta g_{\mu\nu}} (\mathcal{L}_I' + \mathcal{L}_m')\right) = 0. \quad (36)$$

However, the remaining terms involving G in Eq. (36) are very interesting: When one carries out the coordinate transformation to the local geodesic frame, the gravitational interaction does not disappear. The terms involving G represent *a universal short-range gravitational interaction* between the scalar field and other matter.

As a simple example, suppose $\mathcal{L}_I' = 0$ so that there are only gravitational interactions. Suppose further that the only matter present besides the scalar field is a spinor field of mass M. Then (in the geodesic coordinates)

$$\Box^2\phi + \mu^2\phi + \tfrac{1}{3}4\pi G\,\phi(M\bar{\psi}\psi) = 0 \quad (37)$$

which makes the "contact" interaction very obvious. This interaction cannot be eliminated by a counterterm in the Lagrangian. To see this, note that the equation of motion of the spinor field is (also in geodesic coordinates at the same point)

$$-i\gamma^\mu\partial_\mu\psi + M\psi = 0. \quad (38)$$

This equation contains no gravitational terms at all. A counterterm might eliminate the gravitational interaction from Eq. (37); but an extra term will then appear in Eq. (38).

The essential difference between the self-interaction of ϕ and the interaction of ϕ with ψ is that the former behaves as an ordinary nongravitational interaction of special relativity (so much so that, if desired, it can be canceled by a nongravitational counterterm), while the latter is impossible as a direct nongravitational interaction. This impossibility is strikingly demonstrated by Eqs. (37) and (38) which, if interpreted as equations of special relativity (valid throughout all of space–time), lead to the conclusion that ψ acts on ϕ, but not conversely. Only if the gravitational fields are taken into account is this contradiction avoided since then the equations have the special form (37), (38) only at one point. An example of the effects of the new short range interaction on gravitational scattering is given in Appendix B.

Of course, this new interaction is extremely weak. The characteristic dimensionless coupling constant is $(GM^2/\hbar c) \sim 10^{-39}$, the same as for the ordinary long range gravitational interaction. The short range interaction is, therefore, going to be masked very effectively by the other known short range interactions. However, it might possibly be of some importance in the renormalization of gravitational theory.

V. CONCLUSION

The gravitational theory based on the new improved energy–momentum tensor satisfies the weak principle of equivalence. It also satisfies the strong principle of equivalence, even though it is a theory with nonminimal coupling. Thus, although minimal coupling implies the SPE, the converse is not true. This comes as somewhat of a surprise since it has often been argued (e.g., Ref. 10, p. 337) that the SPE requires that in the local geodesic frame the laws of physics must be those of special relativity. We now see [compare Eq. (37)] that in the local geodesic frame the laws can differ from those of special relativity without any violation of the SPE. The nonminimal coupling terms which would usually be expected to produce a violation of the SPE conspire to produce instead a short range gravitational interaction.

APPENDIX A: RELATION BETWEEN ENERGY–MOMENTUM TENSORS

In this appendix we establish a general relation [Eq. (A21)] between the energy–momentum tensor of matter given by Eq. (6) and the corresponding canonical tensor constructed from $\mathcal{L}_m$.

Consider the integral

$$\int (\mathcal{L}_{\min} + \sqrt{-g}R_{\alpha\beta\sigma\rho}f^{\alpha\beta\sigma\rho})\,d^4x. \quad (A1)$$

We can also write this as

$$\int (\mathcal{L}_m + S^\mu{}_{,\mu})\,d^4x, \quad (A2)$$

where the complete divergence $S^\mu{}_{,\mu}$ is the difference between $\sqrt{-g}R_{\alpha\beta\sigma\rho}f^{\alpha\beta\sigma\rho}$ and "$\sqrt{-g}R_{\alpha\beta\sigma\rho}f^{\alpha\beta\sigma\rho}$". [$\mathcal{L}_m$ has been defined in Eq. (15).]

The integrand of (A2) transforms as a scalar density under general coordinate transformations. This means that corresponding to an infinitesimal transformation $x^\mu \to x^\mu + \xi^\mu$, the transformation of the integrand (regarded as an *active* transformation) is

$$\delta(\mathcal{L}_m + S^\mu{}_{,\mu}) = -[(\mathcal{L}_m + S^\rho{}_{,\rho})\xi^\mu]_{,\mu}. \quad (A3)$$

We can combine this with

$$\delta(\mathcal{L}_m) = \left(\delta Q_i \frac{\partial}{\partial Q_i} + \delta Q_{i,\mu} \frac{\partial}{\partial Q_{i,\mu}}\right)\mathcal{L}_m = \delta Q_i \frac{\delta \mathcal{L}_m}{\delta Q_i} + \left[\delta Q_i \frac{\partial \mathcal{L}_m}{\partial Q_{i,\mu}}\right]_{,\mu} \quad (A4)$$

to get Noether's identity (Ref. 10, p. 92)

$$\delta t^{\mu}{}_{,\mu} = \delta Q_i \frac{\delta \mathcal{L}_m}{\delta Q_i}, \tag{A5}$$

where

$$\delta t^{\mu} = -\left((\mathcal{L}_m + S^{\rho}{}_{,\rho})\xi^{\mu} + \delta Q_i \frac{\partial \mathcal{L}_m}{\partial Q_{i,\mu}} + \delta S^{\mu}\right). \tag{A6}$$

The equations of motion of matter are $(\delta/\delta q_i)\mathcal{L}_m = 0$. Hence Eq. (A5) reduced to

$$\delta t^{\mu}{}_{,\mu} = \delta g_{\mu\nu} \frac{\delta \mathcal{L}_m}{\delta g_{\mu\nu}} \tag{A7}$$

which we can also write

$$\delta t^{\mu}{}_{,\mu} = -\tfrac{1}{2}\delta g_{\mu\nu}\tau^{\mu\nu}, \tag{A8}$$

where $\tau^{\mu\nu} = \sqrt{-g}\,T^{\mu\nu}$. The change in $g_{\mu\nu}$ is

$$\delta g_{\mu\nu} = -g_{\mu\alpha}\xi^{\alpha}{}_{,\nu} - g_{\alpha\nu}\xi^{\alpha}{}_{,\mu} - g_{\mu\nu,\alpha}\xi^{\alpha} \tag{A9}$$

which gives

$$\delta t^{\mu}{}_{,\mu} = (\tau^{\mu\nu}g_{\mu\alpha}\xi^{\alpha})_{,\nu} - \sqrt{-g}\,T^{\mu\nu}{}_{;\nu}g_{\mu\alpha}\xi^{\alpha}. \tag{A10}$$

Since $T^{\mu\nu}{}_{;\nu} = 0$, Eq. (A10) reduces to

$$\delta t^{\mu}{}_{,\mu} = (\tau_{\mu}{}^{\nu}\xi^{\mu})_{,\nu}. \tag{A11}$$

We next need an explicit expression for δt^{μ} in terms of ξ^{μ}. The fields Q_l transform according to

$$\delta Q_l = -Q_{l,\mu}\xi^{\mu} - \Lambda_{ln\mu}{}^{\alpha}\xi^{\mu}{}_{,\alpha}Q_n, \tag{A12}$$

where the matrixes $\Lambda_{ln\mu}{}^{\alpha}$ are constant and are completely determined by the Lorentz transformation properties of the fields. The change in S^{μ} can be written

$$\delta S^{\mu} = \xi^{\alpha}\frac{\partial(\delta S^{\mu})}{\partial \xi^{\alpha}} + \xi^{\alpha}{}_{,\beta}\frac{\partial(\delta S^{\mu})}{\partial \xi^{\alpha}{}_{,\beta}} + \xi^{\alpha}{}_{,\beta,\rho}\frac{\partial(\delta S^{\mu})}{\partial \xi^{\alpha}{}_{,\beta,\rho}}. \tag{A13}$$

(Since S^{μ} contains only first derivatives of the fields, δS^{μ} will at most contain second derivatives of ξ^{μ}.) We then obtain

$$-\left[(\mathcal{L}_m + S^{\rho}{}_{,\rho})\xi^{\mu} + \frac{\partial \mathcal{L}_m}{\partial Q_{L,\mu}}(-Q_{l,\alpha}\xi^{\alpha} - \Lambda_{ln\alpha}{}^{\beta}\xi^{\alpha}{}_{,\beta}Q_n) + \xi^{\alpha}\frac{\partial(\delta S^{\mu})}{\partial \xi^{\alpha}} + \xi^{\alpha}{}_{,\beta}\frac{\partial(\delta S^{\mu})}{\partial \xi^{\alpha}{}_{,\beta}} + \xi^{\alpha}{}_{,\beta,\rho}\frac{\partial(\delta S^{\mu})}{\partial \xi^{\alpha}{}_{,\beta,\rho}}\right]_{,\mu} = (\tau_{\mu}{}^{\nu}\xi^{\mu})_{,\nu}. \tag{A14}$$

Equation (A14) must be true for every choice of the functions ξ^{μ}. In particular, we choose ξ^{α} such that at some given point $\xi^{\alpha} = 0$; but the first, second, and third derivatives are not zero. Since $\xi^{\alpha}{}_{,\beta}$ is arbitrary, the coefficients of this quantity on both sides of the above equation must be equal. This implies

$$-(\mathcal{L}_m + S^{\rho}{}_{,\rho})\delta_{\alpha}{}^{\beta} + \frac{\partial \mathcal{L}_m}{\partial Q_{l,\beta}}Q_{l,\alpha} + \left[\frac{\partial \mathcal{L}_m}{\partial Q_{l,\mu}}\Lambda_{ln\alpha}{}^{\beta}Q_n - \frac{\partial(\delta S^{\mu})}{\partial \xi^{\alpha}{}_{,\beta}}\right]_{,\mu} - \frac{\partial(\delta S^{\beta})}{\partial \xi^{\alpha}} = \tau_{\alpha}{}^{\beta}. \tag{A15}$$

The second derivatives are also arbitrary, except for the constraint imposed by the symmetry condition $\xi^{\alpha}{}_{,\beta,\rho} = \xi^{\alpha}{}_{,\rho,\beta}$. Hence the coefficient of $\xi^{\alpha}{}_{,\beta,\rho}$ in Eq. (A14) must either be zero or else equal to some quantity $W_{\alpha}{}^{[\beta\rho]}$ antisymmetric in β,ρ:

$$\frac{\partial \mathcal{L}_m}{\partial Q_{l,\rho}}\Lambda_{ln\alpha}{}^{\beta}Q_n - \frac{\partial(\delta S^{\rho})}{\partial \xi^{\alpha}{}_{,\beta}} - \frac{\partial}{\partial x^{\mu}}\frac{\partial(\delta S^{\mu})}{\partial \xi^{\alpha}{}_{,\beta,\rho}} = W_{\alpha}{}^{[\beta\rho]}. \tag{A16}$$

Finally, we examine the terms involving $\xi^{\alpha}{}_{,\beta,\rho,\sigma}$. This yields the condition

$$\frac{\partial(\delta S^{\sigma})}{\partial \xi^{\alpha}{}_{,\beta,\rho}}\xi^{\alpha}{}_{,\beta,\rho,\sigma} = 0, \tag{A17}$$

which implies that

$$\frac{\partial(\delta S^{\sigma})}{\partial \xi^{\alpha}{}_{,\beta,\rho}} \tag{A18}$$

is antisymmetric in ρ,σ [or, alternatively, antisymmetric in β,σ; these alternatives are equivalent since (A18) is symmetric in β,ρ]. As a consequence of Eq. (A16) we have

$$\left[\frac{\partial \mathcal{L}_m}{\partial Q_{l,\mu}}\Lambda_{ln\alpha}{}^{\beta}Q_n - \frac{\partial(\delta S^{\mu})}{\partial \xi^{\alpha}{}_{,\beta}}\right]_{,\mu} = W_{\alpha}{}^{[\beta\mu]}{}_{,\mu} + \frac{\partial^2}{\partial x^{\mu}\partial x^{\sigma}}\frac{\partial(\delta S^{\sigma})}{\partial \xi^{\alpha}{}_{,\beta,\mu}}. \tag{A19}$$

The second term on the right-hand side vanishes because of the antisymmetry of the expression (A18) in ρ,σ. Furthermore,

$$\frac{\partial(\delta S^{\beta})}{\partial \xi^{\alpha}} = -S^{\beta}{}_{,\alpha}. \tag{A20}$$

Using these results, we find that Eq. (A15) reduces to

$$-\mathcal{L}_m\delta_{\alpha}{}^{\beta} + Q_{l,\alpha}\frac{\partial \mathcal{L}_m}{\partial Q_{l,\beta}} + W_{\alpha}{}^{[\beta\mu]}{}_{,\mu} - S^{\rho}{}_{,\rho}\delta_{\alpha}{}^{\beta} + S^{\beta}{}_{,\alpha} = \tau_{\alpha}{}^{\beta}. \tag{A21}$$

The term involving S^{μ}, can be put in the form

$$[-S^{\rho}\delta_{\alpha}{}^{\beta} + \delta_{\alpha}{}^{\rho}S^{\beta}]_{,\rho}, \tag{A22}$$

where the tensor in the brackets is obviously antisymmetric in ρ,β. Hence the general relation (A21) simply states that the canonical tensor constructed from $\mathcal{L}_m$ and the tensor $\tau_{\alpha}{}^{\beta}$ constructed from Eq. (6) differ only by a term of the form $U_{\alpha}{}^{[\beta\mu]}{}_{,\mu}$. If we assume that the matter is confined to a finite volume, the space integrals of the ${}_0{}^{\alpha}$ components of these tensors will then be equal, as was claimed in Eq. (25). This result does not depend on the detailed form of $\mathcal{L}_m$ and S^{μ}. The only essential requirement is that $\mathcal{L}_m + S^{\mu}{}_{,\mu}$ transform as a scalar density and that $\mathcal{L}_m$ and S^{μ} contain only first derivatives and these at most quadratically.

APPENDIX B: GRAVITATIONAL SCATTERING

The following is an example of the effects of the short range interaction on the gravitational scattering, according to quantum theory, of a spin zero particle by a spin-$\frac{1}{2}$ particle. This example is similar to one given in Ref. 2.

For a transition $k \to k', p \to p'$ (where k and p stand, respectively, for the boson and fermion momenta), the transition matrix element to lowest order in G is proportional to

$$4\pi G\bar{u}_{s'}(p')\tfrac{1}{2}(p'^{\mu} + p^{\mu})\gamma^{\nu}u_s(p)D_{\mu\nu\alpha\beta}(q) \times (1/2\mu)[2k'^{\alpha}k^{\beta} - \eta^{\alpha\beta}(k\cdot k' - \mu^2) - \tfrac{1}{3}(q^{\alpha}q^{\beta} - \eta^{\alpha\beta}q^2)], \tag{B1}$$

where

$$D_{\mu\nu\alpha\beta}(q) = (1/q^2)(\eta_{\mu\alpha}\eta_{\nu\beta} + \eta_{\mu\beta}\eta_{\nu\alpha} - \eta_{\mu\nu}\eta_{\alpha\beta}) \quad \text{(B2)}$$

is the graviton propagator and $q = k' - k$. To first approximation, the fermion energy–momentum tensor is conserved and (B1) simplifies to

$$4\pi G\bar{u}_{s'}(p')\tfrac{1}{2}(p'^{\mu} + p^{\mu})\gamma^{\nu}u_s(p) \times (1/2\mu)\{D_{\mu\nu\alpha\beta}(q)[2k'^{\alpha}k^{\beta} - \eta^{\alpha\beta}(k\cdot k' - \mu^2)] - \tfrac{1}{3}\eta_{\mu\nu}\}. \quad \text{(B3)}$$

If we proceed to the nonrelativistic limit and assume that M is large compared to μ, we obtain

$$\bar{u}_s(p)u_{s'}(p')[(4\pi GM/-\mathbf{q}^2) - (2\pi GM/3\mu)]. \quad \text{(B4)}$$

The first term corresponds to scattering by the familiar potential

$$-GM\mu/r. \quad \text{(B5)}$$

The second term behaves as though due to a potential

$$-(2\pi GM/3\mu)\delta^3(\mathbf{r}), \quad \text{(B6)}$$

i.e., the gravitational interaction between spin-zero and spin-$\frac{1}{2}$ particles seems to have a hard attractive core.

[1]We take $c = 1$ and a metric of signature + - - -. The Lorentz metric will be designated by $\eta_{\mu\nu}$. The bracket $[\alpha\beta]$ on the indices of a tensor will indicate antisymmetry.

[2]C. G. Callan, S. Coleman, and R. Jackiw, Ann. Phys. **59**, 42 (1970).

[3]By *general Einstein theory* we shall mean a theory obtained from a Lagrangian containing first derivatives at most quadratically (second derivatives linearly). The Lagrangian is assumed to be invariant under general coordinate transformations and this invariance is achieved by the use of a gauge field $g_{\mu\nu}$. Theories with two gravitational gauge fields or with a long range gravitational scalar field are excluded.

[4]If there are spin matrices γ^{μ} in the Lagrangian, then these must be replaced by matrix fields $\gamma^{\mu}(x)$.

[5]L. Kannenberg and R. Arnowitt, Ann. Phys. **45**, 416 (1967).

[6]R. H. Dicke, "The Observational Basis of General Relativity", in *Gravitation and Relativity*, edited by H.-Y. Chiu and W. F. Hoffmann (Benjamin, New York, 1964).

[7]By *definition*, $g_{\mu\nu}$ is such that the standard test body moves along the geodesics. This means that $g_{\mu\nu}$ is the metric as measured by a Marzke-Wheeler geodesic clock [R. F. Marzke and J. A. Wheeler, "Gravitation as Geometry–I," in *Gravitation and Relativity*, edited by H.-Y. Chiu and W. F. Hoffmann (Benjamin, New York, 1964)] constructed out of identical copies of the standard test body.

[8]C. Brans and R. H. Dicke, Phys. Rev. **124**, 925 (1961).

[9]H. C. Ohanian, Ann. Phys. **67**, 648 (1971).

[10]J. L. Anderson, *Principles of Relativity Physics* (Academic, New York, 1967).

Scintillations of randomized electromagnetic fields. II

Leonard S. Taylor

Electrical Engineering Department, University of Maryland, College Park, Maryland

U. S. Naval Ordnance Laboratory, White Oak, Silver Spring, Maryland

Danai Lekhyananda

Electrical Engineering Department, University of Maryland, College Park, Maryland

Department of Aviation, Ministry of Communications, Bangkok, Thailand

(Received 24 October 1972)

We study the propagation of an electromagnetic wave which has suffered changes of phase and amplitude in its passage through a randomizing medium such as a turbulent dielectric. General formulas are derived for the nth-order autocorrelation and spectrum of intensity fluctuations in terms of the $2n$th-order mutual coherence function on the initial plane. Simple expressions are obtained when the wave is initially weakly phase distributed. The probability distribution of intensity fluctuations is shown to be Gaussian in the limit of vanishing phase perturbation.

Phenomena such as the twinkling of starlight and interplanetary and interstellar radio star scintillations involve the propagation of an electromagnetic wave in uniform media after the wave has suffered random changes of phase and amplitude in its passage through an irregular refracting region. Although the second and fourth moments of such randomized fields are adequate to analyze experiments involving interference and diffraction and to determine the autocorrelation of intensity fluctuations, it is necessary to use higher-order moments to completely characterize the statistical properties of the field and to completely interpret experiments in which intensity fluctuations are measured. In this paper we study the higher-order moments, using techniques developed in previous work[1,2] to obtain general expressions for the nth-order moments of the field and the corresponding spectra. Analytical results obtained by previous authors[3-11] have generally been limited to the second and fourth moments, an exception being the work of Mercier,[4] who obtained the nth moment in the special case of an infinitely distant phase screen with Gaussian autocorrelation. The general formulas derived in this paper are valid at all distances for waves which are initially random in both phase and amplitude, but special results shall be obtained for waves which have passed through such phase screens which introduce phase fluctuations without appreciable disturbance of the wave amplitude. That is, upon entering the uniform medium at $z = 0$, the wave $\psi = Ae^{jS}$ has uniform amplitude and random phase across the x–y (transverse) plane. As the wavefront propagates further from the screen, however, fluctuations of amplitude begin to develop which, of course, are randomly distributed across a plane parallel to the initial plane. Because we are treating the propagation of an electromagnetic field in a uniform isotropic medium, we can employ the scalar wave equation to determine the field in this region. We do not consider the problem of the wave in the nonuniform medium.

1. GENERAL THEORY

We consider the propagation of an electromagnetic wave $\psi(\mathbf{r})$ obeying the scalar wave equation

$$(\nabla^2 + k^2)\psi(\mathbf{r}) = 0. \tag{1}$$

As usual, $\psi(\mathbf{r})$ represents a monochromatic wave with $k^2 = \epsilon k_0^2 = \epsilon(\omega/c)^2$. We shall assume that the boundary values are given over the $z = 0$ plane. The use of Green's theorem leads to the Helmholtz–Kirchhoff integral, which can be expressed in the form[1]

$$\psi(\mathbf{r}) = -\frac{1}{2\pi}\frac{\partial}{\partial z}\iint \psi(\mathbf{r}')\,\frac{e^{jkR}}{R}\,dS', \tag{2}$$

where $R = [(x-x')^2 + (y-y')^2 + z^2]^{1/2}$. Using Eq. (2), we now obtain the following expression for the $2n$th-order mutual coherence function, $M(\mathbf{r}_1, \mathbf{r}_2, \ldots, \mathbf{r}_{2n}) = \langle \psi(\mathbf{r}_1)\psi^*(\mathbf{r}_2)\psi(\mathbf{r}_3)\psi^*(\mathbf{r}_4)\cdots\psi(\mathbf{r}_{2n-1})\psi^*(\mathbf{r}_{2n})\rangle$ in the space $z > 0$ in terms of the $2n$th-order mutual coherence on the boundary plane:

$$\begin{aligned} M(\mathbf{r}_1, \mathbf{r}_2, \ldots, \mathbf{r}_{2n}) &= \left(\frac{1}{2\pi}\right)^{2n} \frac{\partial}{\partial z_1}\cdots\frac{\partial}{\partial z_{2n}} \int_{P_1}\int_{P_2}\cdots\int_{P_{2n}} dS_1'\cdots dS_{2n}' \\ &\times M_0(\mathbf{r}_1', \mathbf{r}_2', \ldots, \mathbf{r}_{2n}') \\ &\times \frac{\exp[jk(R_1 - R_2 + R_3 - R_4 + \cdots + R_{2n-1} - R_{2n})]}{R_1 R_2 \cdots R_{2n}}. \end{aligned} \tag{3}$$

The zero subscript has been placed on the function M inside the integral as a reminder that it is being evaluated over the $z = 0$ plane. We shall assume that the mutual coherence is homogeneous on the boundary plane; it may be written as a function of the difference between coordinates:

$$\begin{aligned} &\xi_1' = x_1' - x_2', \quad \xi_2' = x_2' - x_3', \ldots, \xi_{2n-1}' = x_{2n-1}' - x_{2n}' \\ &\xi_{2n}' = y_1' - y_2', \quad \xi_{2n+1}' = y_2' - y_3', \ldots, \xi_{2(2n-1)}' \\ &\qquad = y_{2n-1}' - y_{2n}', \end{aligned} \tag{4}$$

and $\boldsymbol{\xi}'$ is a vector in $2(2n-1)$-space. It should be noted that the differences $x_{2n}' - x_1'$ and $y_{2n}' - y_1'$ can be expressed in terms of the other differences and have not been included explicitly. Any other differences, such as $x_i' - x_j'$, where i and j are two different integers with $i < j$, can be calculated, for example, from

$$x_i' - x_j' = \xi_i' + \xi_{i+1}' + \cdots + \xi_{j-1}'. \tag{5}$$

We now introduce the $2(2n-1)$-fold Fourier transform $\hat{M}_0(\mathbf{k}_\xi)$,

$$M_0(\boldsymbol{\xi}') = (1/2\pi)^{2(2n-1)} \int^{(2[2n-1])} d\mathbf{k}_\xi \exp(j\mathbf{k}_\xi \cdot \boldsymbol{\xi}')\hat{M}_0(\mathbf{k}_\xi), \tag{6}$$

where $k_{\xi_1}, \ldots, k_{\xi_{2(2n-1)}}$ are the spatial frequencies corresponding to $\xi'_1, \ldots, \xi'_{2(2n-1)}$ and $d\mathbf{k}_\xi = dk_{\xi_1} \cdots dk_{\xi_{2(2n-1)}}$. The integration is $2(2n-1)$-fold, as indicated by the numeral in parentheses, and each integral is taken from $-\infty$ to ∞. We employ this convention throughout unless otherwise indicated. The function $\hat{M}_0(\mathbf{k}_\xi)$ is subject to no restriction except that $\int^{(2[2n-1])} |\hat{M}_0(\mathbf{k}_\xi)|\, d\mathbf{k}_\xi < \infty$. Substituting, we obtain, upon interchanging the order of integration,

$$M(\mathbf{r}_1, \ldots, \mathbf{r}_{2n}) = \left(\frac{1}{2\pi}\right)^{2(3n-1)} \int^{(2[2n-1])} d\mathbf{k}_\xi \hat{M}_0(\mathbf{k}_\xi) \frac{\partial}{\partial z_1} \cdots \frac{\partial}{\partial z_{2n}} \times \int^{(4n)} \frac{\exp[j\mathbf{k}_\xi \cdot \xi' + jk(R_1 - R_2 + \cdots + R_{2n-1} - R_{2n})]}{R_1 R_2 \cdots R_{2n}} \times dx'_1 \cdots dy'_{2n}. \quad (7)$$

By expressing the ξ' in terms of x' and y' and regrouping so that each term in the exponential appears as a factor of x_j or y_j, Eq. (7) may be expressed in the form

$$\begin{aligned} M(\mathbf{r}_1, \ldots, \mathbf{r}_{2n}) &= \left(\frac{1}{2\pi}\right)^{2(2n-1)} \int^{(2[2n-1])} d\mathbf{k}_\xi \hat{M}_0(\mathbf{k}_\xi) \left(\frac{\partial}{\partial z_1} T(\mathbf{r}_1; k_{\xi_1}, k_{\xi_{2n}})\right) \\ &\times \left(\frac{\partial}{\partial z_2} T^*(\mathbf{r}_2; k_{\xi_1} - k_{\xi_2}, k_{\xi_{2n}} - k_{\xi_{2n+1}})\right) \\ &\times \left(\frac{\partial}{\partial z_3} T(\mathbf{r}_3; k_{\xi_3} - k_{\xi_2}, k_{\xi_{2(n+1)}} - k_{\xi_{2n+1}})\right) \\ &\times \left(\frac{\partial}{\partial z_4} T^*(\mathbf{r}_4; k_{\xi_3} - k_{\xi_4}, k_{\xi_{2(n+1)}} - k_{\xi_{2n+3}})\right) \cdots \\ &\times \left(\frac{\partial}{\partial z_{2(n-1)}} T^*(\mathbf{r}_{2(n-1)}; k_{\xi_{2n-3}} - k_{\xi_{2(n-1)}}, k_{\xi_{4(n-1)}} - k_{\xi_{4n-3}})\right) \\ &\times \left(\frac{\partial}{\partial z_{2n-1}} T(\mathbf{r}_{2n-1}; k_{\xi_{2n-1}} - k_{\xi_{2(n-1)}}, k_{\xi_{2(2n-1)}} - k_{\xi_{4n-3}})\right) \\ &\times \left(\frac{\partial}{\partial z_{2n}} T^*(\mathbf{r}_{2n}; k_{2n-1}, k_{2(2n-1)})\right), \end{aligned} \quad (8)$$

where

$$T(\mathbf{r}_i; k_{\xi_m}, k_{\xi_n}) = \frac{1}{2\pi} \int^{(2)} dx'_i dy'_i \frac{\exp[j(k_{\xi_m} x'_i + k_{\xi_n} y'_i + kR_i)]}{R_i}. \quad (9)$$

Using the formula obtained previously,[2]

$$\frac{\partial}{\partial z_i} T(\mathbf{r}_i; k_{\xi_m}, k_{\xi_n}) = -\exp\{j[k_{\xi_m} x_i + k_{\xi_n} y_i - z_i(k^2 - k^2_{\xi_m} - k^2_{\xi_n})^{1/2}]\}. \quad (10)$$

Next, substituting Eq. (10) into Eq. (8), rearranging the argument of the exponential into factors of k_{ξ_j}, and again expressing x_i and y_i in terms of ξ, we obtain, after additional manipulations,

$$\begin{aligned} M(\mathbf{r}_1, \ldots, \mathbf{r}_{2n}) = \left(\frac{1}{2\pi}\right)^{2(2n-1)} \int^{(2[2n-1])} d\mathbf{k}_\xi \hat{M}_0(\mathbf{k}_\xi) \\ \times \exp(j\mathbf{k}_\xi \cdot \xi + jkz_1[1 - k^{-2}(k^2_{\xi_1} + k^2_{\xi_{2n}})]^{1/2} \\ - jkz_2\{1 - k^{-2}[(k_{\xi_1} - k_{\xi_2})^2 + (k_{\xi_{2n}} - k_{\xi_{2n+1}})^2]\}^{1/2} \\ + jkz_3\{1 - k^{-2}[(k_{\xi_3} - k_{\xi_2})^2 + (k_{\xi_{2(n+1)}} - k_{\xi_{2n+1}})^2]\}^{1/2} \\ - \cdots - jkz_{2(n-1)}\{1 - k^2[(k_{\xi_{2n-3}} - k_{\xi_{2(n-1)}})^2 \\ + (k_{\xi_{4(n-1)}} - k_{\xi_{4n-3}})^2]\}^{1/2} \\ + jkz_{2n-1}\{1 - k^2[(k_{\xi_{2n-1}} - k_{\xi_{2(n-1)}})^2 \\ + (k_{\xi_{2(2n-1)}} - k_{\xi_{4n-3}})^2]\}^{1/2} \\ - jkz_{2n}[1 - k^{-2}(k^2_{\xi_{2n-1}} + k^2_{\xi_{2(2n-1)}})]^{1/2}). \end{aligned} \quad (11)$$

We can assume that $\hat{M}_0$ is nonzero only for k_ξ such that each of the square roots appearing in Eq. (11) is real. This restriction on the spectrum can be interpreted in terms of the angular spectrum of plane waves to mean that evanescent waves are not included: These waves would rapidly decay in any case. In order to evaluate the integral, we shall also expand each of the square roots in series and retain only the first two terms of each expansion. If l is the transverse scale of the wave fluctuations on the initial plane, it is readily seen that these approximations require that $l > \lambda = 2\pi/k$ and $z\lambda^3/l^4 \ll 1$. These requirements appear to be readily satisfied for all cases involving waves which have been randomized by terrestrial and solar atmospheric effects. For example, in the case of stellar optical scintillations due to the upper atmosphere, z is approximately 10^4 m, $\lambda \sim 10^{-7}$ m, and $l \sim 10^2$; then $z\lambda^3/l^4 = 10^{-25}$. For radio star scintillations due to the ionosphere, the typical values are $l \sim 10^5$ m, $\lambda \sim 10$ m, $l \sim 10^3$ m; then $z\lambda^3/l^4 = 10^{-4}$. For radio star scintillations due to the solar plasma, z is approximately 10^{11} m, $\lambda \sim 1$ m, and l has the dimension comparable to the Sun $\sim 10^9$ m; then $z\lambda^3/l^4 = 10^{-25}$.

Proceeding with the expansions and limiting ourselves to the important case of the transverse coherence function by taking $z_1 = z_2 = \cdots = z_{2n} = z$, we obtain after some simplification

$$\begin{aligned} M(\mathbf{r}_1, \ldots, \mathbf{r}_{2n}) &= \left(\frac{1}{2\pi}\right)^{2(2n-1)} \int^{(2[2n-1])} d\mathbf{k}_\xi \hat{M}_0(\mathbf{k}_\xi) \\ &\times \exp\{j\mathbf{k}_\xi \cdot \xi + (jz/k)[(k_{\xi_1} k_{\xi_2} - k_{\xi_2} k_{\xi_3} + k_{\xi_3} k_{\xi_4} - \cdots \\ &- k_{\xi_{2(n-1)}} k_{\xi_{2n-1}}) + (k_{\xi_{2n}} k_{\xi_{2n+1}} - k_{\xi_{2n+1}} k_{\xi_{2(n+1)}} \\ &+ \cdots - k_{\xi_{4n-3}} k_{\xi_{2(2n-1)}})]\} \\ &= \mathfrak{F}^{-1}_{2[2n-1]}\{\hat{M}_0(\mathbf{k}_\xi) h_x h_y\} \\ &= M_0(\xi) \bigstar \mathfrak{F}^{-1}_{2n-1}\{h_x\}\, \mathfrak{F}^{-1}_{2n-1}\{h_y\}, \end{aligned} \quad (12)$$

where $\mathfrak{F}^{-1}_{2[2n-1]}$ is the inverse transform of order $2(2n-1)$ and the star indicates the convolution. We have defined

$$h_x = \exp[(jz/k)(k_{\xi_1} k_{\xi_2} - k_{\xi_2} k_{\xi_3} + k_{\xi_3} k_{\xi_4} - \cdots - k_{\xi_{2(n-1)}} k_{\xi_{2n-1}})],$$

$$h_y = \exp[(jz/k)(k_{\xi_{2n}} k_{\xi_{2n+1}} - k_{\xi_{2n+1}} k_{\xi_{2(n+1)}} + k_{\xi_{2(n+1)}} k_{\xi_{2n+3}} - \cdots - k_{\xi_{4n-3}} k_{\xi_{2(2n-1)}})]. \quad (13)$$

The inverse transform, $\mathfrak{F}^{-1}_{2n-1}\{h_x\}$, after rearrangement of terms, is given by

$$\mathfrak{F}^{-1}_{2n-1}\{h_x\} = (1/2\pi)^{(2n-1)} \int^{(2n-1)} dk_{\xi_1} \cdots dk_{\xi_{2n-1}}$$

$$\times \exp(j\{k_{\xi_2}[(z/k)(k_{\xi_1}-k_{\xi_3})+\xi_2] + k_{\xi_4}[(z/k)(k_{\xi_3}-k_{\xi_5})+\xi_4]+\cdots + k_{\xi_{2(n-1)}}[(z/k)(k_{\xi_{2n-3}}-k_{\xi_{2n-1}})+\xi_{2(n-1)}]\}) \times \exp\{j(k_{\xi_1}\xi_1 + k_{\xi_3}\xi_3 + k_{\xi_5}\xi_5 + \cdots + k_{\xi_{2n-1}}\xi_{2n-1})]. + k_{\xi_5} + \cdots + k\xi_{2n-1})]. \quad (14)$$

Integrating over $k_{\xi_2}, k_{\xi_4}, \ldots, k_{\xi_{2(n-1)}}$, we obtain

$$\mathfrak{F}^{-1}_{2n-1}\{h_x\} = (1/2\pi)^n \int^{(n)} dk_{\xi_1} dk_{\xi_3} \cdots dk_{\xi_{2n-1}} \times \delta[(z/k)(k_{\xi_1}-k_{\xi_3})+\xi_2] \times \delta[(z/k)(k_{\xi_3}-k_{\xi_5})+\xi_4] \times\cdots\delta[(z/k)(k_{\xi_{2n-3}}-k_{\xi_{2n-1}})-\xi_{2(n-1)}] \times \exp[j(k_{\xi_1}\xi_1 + k_{\xi_3}\xi_3 + \cdots + k_{\xi_{2n-1}}\xi_{2n-1})], \quad (15)$$

where δ is the Dirac delta function. It is convenient to introduce the transformations

$$\alpha_{n-1} = (z/k)(k_{\xi_{2n-3}} - k_{\xi_{2n-1}}) + \xi_{2(n-1)}, \qquad \beta = \tfrac{1}{2}(k_{\xi_1} + k_{\xi_{2n-1}}). \quad (16)$$

The inverse transformations are

$$k_{\xi_1} = \beta + (k/2z)(\alpha_1 + \alpha_2 + \cdots + \alpha_{n-1} - \xi_2 - \xi_4 - \cdots - \xi_{2(n-1)}),$$

$$k_{\xi_3} = \beta + (k/2z)(-\alpha_1 + \alpha_2 + \cdots + \alpha_{n-1} + \xi_2 - \xi_4 - \cdots - \xi_{2(n-1)}),$$

$$k_{\xi_{2n-3}} = \beta + (k/2z)(-\alpha_1 - \alpha_2 - \cdots - \alpha_{n-2} + \alpha_{n-1} + \xi_2 + \cdots + \xi_{2(n-2)} - \xi_{2(n-1)}),$$

$$k_{\xi_{2n-1}} = \beta + (k/2z)(-\alpha_1 - \cdots - \alpha_{n-1} + \xi_2 + \xi_4 + \cdots + \xi_{2(n-1)}). \quad (17)$$

The Jacobian of the transformation can be found from the $n \times n$ determinant:

$$\begin{vmatrix} k/2z & k/2z & \cdots & k/2z & 1 \\ -k/2z & k/2z & \cdots & k/2z & 1 \\ -k/2z & -k/2z & \cdots & k/2z & 1 \\ \cdot & & & & \cdot \\ \cdot & & & & \\ \cdot & & & & \\ -k/2z & -k/2z & \cdots & -k/2z & 1 \end{vmatrix}. \quad (18)$$

Factoring the common factor, $k/2z$, out of the first $(n-1)$ columns, the remaining $n \times n$ determinant with all elements below the diagonal equal to -1 and the rest equal to 1 is found to be equal to $(2)^{n-1}$. Thus the Jacobian is $(2)^{n-1}(k/2z)^{n-1} = (k/z)^{n-1}$:

$$\mathfrak{F}^{-1}_{2n-1}\{h_x\} = (1/2\pi)^n (k/z)^{n-1} \int^{(n)} d\alpha_1 d\alpha_2 \cdots d\alpha_{n-1} d\beta \times \delta(\alpha_1)\,\delta(\alpha_2)\,\delta(\alpha_3)\cdots\delta(\alpha_{n-1}) \times \exp(j\{\xi_1[\beta + (k/2z)(\alpha_1+\alpha_2+\cdots+\alpha_{n-1}-\xi_2 - \xi_4 - \cdots - \xi_{2(n-1)})] + \xi_3[\beta + (k/2z)(-\alpha_1+\alpha_2+\cdots+\alpha_{n-1}+\xi_2 - \xi_4 - \cdots - \xi_{2(n-1)})] + \cdots + \xi_{2n-1}[\beta + (k/2z)(-\alpha_1-\alpha_2-\cdots - \xi_4 - \cdots - \xi_{2(n-1)})] + \cdots + \xi_{2n-1}[\beta + (k/2z)(-\alpha_1-\alpha_2-\cdots - \alpha_{n-1}+\xi_2+\xi_4+\cdots+\xi_{2(n-1)})]\}) = (k/2\pi z)^{n-1}\,\delta(\xi_1+\xi_3+\xi_5+\cdots+\xi_{2n-1}) \times \exp\{(jk/2z)[\xi_2(-\xi_1+\xi_3+\xi_5+\cdots+\xi_{2n-1} + \xi_4(-\xi_1-\xi_3+\xi_5+\cdots+\xi_{2n-1}) + \cdots + \xi_{2(n-1)}(-\xi_1-\xi_3-\xi_5-\cdots+\xi_{2n-1})]\}, \quad (19)$$

the integration over β leading to the delta function term. Similarly,

$$\mathfrak{F}^{-1}_{2n-1}\{h_y\} = (k/2\pi z)^{n-1}\,\delta(\xi_{2n} + \xi_{2(n+1)} + \xi_{2(n+2)} + \cdots + \xi_{2(2n-1)}) \times \exp\{(jk/2z)[\xi_{2n+1}(-\xi_{2n} + \xi_{2(n+1)} + \xi_{2(n+2)} + \cdots + \xi_{2(2n-1)}) + \xi_{2n+3}(-\xi_{2n} - \xi_{2(n+1)} + \cdots + \xi_{2(2n-1)}) + \cdots + \xi_{4n-3}(-\xi_{2n} - \xi_{2(n+1)} - \xi_{2(n+2)} - \cdots - \xi_{4(n-1)} + \xi_{2(2n-1)})]\}. \quad (20)$$

Thus the expression for the $2n$th-order mutual coherence function in Eq. (12) can be written as

$$M(\xi, z) = M_0(\xi) \star (k/2\pi z)^{2(n-1)}\,\delta(\xi_1+\xi_3+\xi_5+\cdots+\xi_{2n-1}) \times \delta(\xi_{2n} + \xi_{2(n+1)} + \xi_{2(n+2)} + \cdots + \xi_{2(2n-1)}) \times \exp\{(jk/2z)[\xi_2(-\xi_1+\xi_3+\xi_5+\cdots+\xi_{2n-1}) + \xi_4(-\xi_1-\xi_3+\xi_5+\cdots+\xi_{2n-1}) + \cdots + \xi_{2(n-1)}(-\xi_1-\xi_3-\xi_5-\cdots-\xi_{2n-3}+\xi_{2n-1}) + \xi_{2n+1}(-\xi_{2n}+\xi_{2(n+1)}+\xi_{2(n+2)}+\cdots+\xi_{2(2n-1)}) + \xi_{2n+3}(-\xi_{2n}-\xi_{2(n+1)}+\xi_{2(n+2)}+\cdots+\xi_{2(2n-1)} + \cdots + \xi_{4n-3}(-\xi_{2n}-\xi_{2(n+1)} - \cdots - \xi_{4(n-1)} + \xi_{2(2n-1)})]\}. \quad (21)$$

The quantity of primary interest is the nth-order autocorrelation of intensity $\langle I(p_1,q_1,z)I(p_2,q_2,z)I(p_3,q_3,z)\cdots I(p_{n-1},q_{n-1},z)I(0,0,z)\rangle$. Thus, taking $x_1 = x_2$, $x_3 = x_4$, etc., we set $\xi_1 = \xi_3 = \xi_5 = \cdots = \xi_{2n-1} = \xi_{2n} = \xi_{2(n+1)} = \cdots = \xi_{2(2n-1)} = 0$; also $\xi_2 = p_1, \xi_4 = p_2, \ldots, \xi_{2(n-1)} = p_{n-1}$, and similarly $\xi_{2n+1} = q_1, \xi_{2n+3} = q_2, \ldots, \xi_{4n-1} = q_{n-1}$. We obtain

$$\langle I(p_1,q_1,z)I(p_2,q_2,z)\cdots I(p_{n-1},q_{n-1},z)I(0,0,z)\rangle = (k/2\pi z)^{2(n-1)} \int^{(2[2n-1])} d\xi' M_0(\xi') \times \delta(-\xi'_1 - \xi'_3 - \xi'_5 - \cdots - \xi'_{2n-1}) \times \delta(-\xi'_{2n} - \xi'_{2(n+1)} - \xi'_{2(n+2)} - \cdots - \xi'_{2(2n-1)}) \exp\{(jk/2z)[(p_1-\xi'_2) \times (\xi'_1 - \xi'_3 - \xi'_5 - \cdots - \xi'_{2n-1}) + (p_2 - \xi'_4)(\xi'_1 + \xi'_3 - \xi'_5 - \cdots - \xi'_{2n-1}) + \cdots + (p_{n-1} - \xi'_{2(n-1)})(\xi'_1 + \xi'_3 + \xi'_5 + \cdots + \xi'_{2n-3} - \xi'_{2n-1}) + (q_1 - \xi'_{2n+1})(\xi'_{2n} - \xi'_{2(n+1)} - \xi'_{2(n+2)} - \cdots - \xi'_{2(2n-1)}) + (q_2 - \xi'_{2n+3}) \times (\xi'_{2n} + \xi'_{2(n+1)} - \xi'_{2(n+2)} - \cdots - \xi'_{2(2n-1)}) + \cdots + (q_{n-1} - \xi'_{4n-3}) \times (\xi'_{2n} + \xi'_{2(n+1)} + \cdots + \xi'_{4(n-1)} - \xi'_{2(2n-1)})]\}. \quad (22)$$

Integrating over $\xi_{2n-1}, \xi_{2(2n-1)}$, we have

$$\begin{aligned}
&\langle I(p_1,q_1,z)\cdots I(0,0,z)\rangle \\
&\quad = (k/2\pi z)^{2(n-1)} \int^{(4[n-1])} d\xi_1' d\xi_2' \cdots d\xi_{2(n-1)}' \\
&\quad\times d\xi_{2n}' d\xi_{2n+1}' \cdots d\xi_{4n-3}' M_0[\xi_1', \xi_2', \xi_3', \ldots, \xi_{2(n-1)}', \\
&\quad - (\xi_1' + \xi_3' + \xi_5' + \cdots + \xi_{2n-3}'), \xi_{2n}', \xi_{2n+1}', \ldots, \xi_{4n-3}', \\
&\quad - (\xi_{2n}' + \xi_{2(n+1)}' + \cdots + \xi_{4(n-1)}')] \\
&\quad\times \exp\{(jk/z)[(p_1 - \xi_2')\xi_1' + (p_2 - \xi_4')(\xi_1' + \xi_3') \\
&\quad + (p_3 - \xi_6')(\xi_1' + \xi_3' + \xi_5') + \cdots + (p_{n-1} - \xi_{2(n-1)}') \\
&\quad\times (\xi_1' + \xi_3' + \xi_5' + \cdots + \xi_{2n-3}') \\
&\quad + (q_1 - \xi_{2n+1}')\xi_{2n}' + (q_2 - \xi_{2n+3}') \\
&\quad\times (\xi_{2n}' + \xi_{2(n+1)}') + \cdots + (q_{n-1} - \xi_{4n-3}') \\
&\quad\times (\xi_{2n}' + \xi_{2(n+1)}' + \xi_{2(n+2)}' + \cdots + \xi_{4(n-1)}')]\}. \qquad (23)
\end{aligned}$$

Transforming by setting $p_i - \xi_{2i}' = \xi_{2i}$ and $q_i - \xi_{2(n+i)-1}' = \xi_{2(n+i)-1}$, where $i = 1, 2, \ldots, n-1$, and also dropping the primes on the ξ_i' and ξ_{2n+i-1}', where i is odd, we have

$$\begin{aligned}
&\langle I(p_1,q_1,z)\cdots I(0,0,z)\rangle \\
&\quad = \left(\frac{k}{2\pi z}\right)^{2(n-1)} \int^{(4[n-1])} d\xi_1 \cdots d\xi_{2(n-1)} \\
&\quad\times d\xi_{2n} d\xi_{2n+1} \cdots d\xi_{4n-3} M_0[\xi_1, (p_1 - \xi_2), \xi_3, \\
&\quad (p_2 - \xi_4), \ldots, (p_{n-1} - \xi_{2(n-1)}), \\
&\quad - (\xi_1 + \xi_3 + \xi_5 + \ldots + \xi_{2n-3}), \xi_{2n}, (q_1 - \xi_{2n+1}), \\
&\quad \ldots, (q_{n-1} - \xi_{4n-3}), - (\xi_{2n} + \xi_{2(n+1)} + \xi_{2(n+2)} \\
&\quad + \ldots + \xi_{4(n-1)})] \exp\left[\frac{jk}{z}\left(\sum_{i=1}^{n-1}\sum_{m=0}^{i-1} \xi_{2i}\xi_{2m+1}\right.\right. \\
&\quad \left.\left. + \sum_{i=1}^{n-1}\sum_{m=0}^{i-1} \xi_{2(n+i)-1}\xi_{2(m+n)}\right)\right]. \qquad (24)
\end{aligned}$$

The nth moment of the intensity field $\langle I^n(z)\rangle$ is obtained from Eq. (24) by setting $p_i, q_i = 0$. Experimental investigations and previous theoretical work in radio astronomy have been particularly concerned with the spectrum of intensity fluctuations. The definition for the nth-order generalization is

$$\begin{aligned}
&\hat{M}_{nI}(f_{x_1}, f_{x_2}, \ldots, f_{x_{n-1}}, f_{y_1}, f_{y_2}, \ldots, f_{y_{n-1}}; z) \\
&\quad = \mathfrak{F}_{p_1,p_2,\ldots,p_{n-1},q_1,q_2,\ldots,q_{n-1}}[\langle I(p_1,q_1,z) \\
&\quad\times I(p_2,q_2,z)\cdots I(p_{n-1},q_{n-1},z) I(0,0,z)\rangle - 1], \qquad (25)
\end{aligned}$$

where $\mathfrak{F}_{p_1,p_2,\ldots,p_{n-1},q_1,q_2,\ldots,q_{n-1}}$ indicates the Fourier transform with respect to $p_1, p_2, \ldots, p_{n-1}, q_1, q_2, \ldots, q_{n-1}$. We restrict ourselves to presenting the result obtained upon substituting (24) into (25) and carrying through a number of manipulations. We find

$$\begin{aligned}
\hat{M}_{nI} &= \mathfrak{F}_{\alpha_1,\alpha_2,\ldots,\alpha_{n-1},\beta_1,\beta_2,\ldots,\beta_{n-1}}\{M_0[zf_{x_1}/k, \alpha_1, \\
&\quad z(f_{x_2} - f_{x_1})/k, \alpha_2, z(f_{x_3} - f_{x_2})/k, \ldots, \\
&\quad \alpha_{n-2}, z(f_{x_{n-1}} - f_{x_{n-2}})/k, \alpha_{n-1}, \\
&\quad - zf_{x_{n-1}}/k, zf_{y_1}/k, \beta_1, z(f_{y_2} - f_{y_1})/k, \ldots, \\
&\quad z(f_{y_{n-1}} - f_{y_{n-2}})/k, \beta_{n-1}, - zf_{y_{n-1}}/k] - 1\}. \qquad (26)
\end{aligned}$$

Equation (26) may be written in terms of the original variables in the form

$$\begin{aligned}
\hat{M}_{nI} &= \mathfrak{F}_{\alpha_1,\alpha_2,\ldots,\alpha_{n-1},\beta_1,\beta_2,\ldots,\beta_{n-1}}\{M_0[x_1 - x_2 = zf_{x_1}/k, \\
&\quad x_2 - x_3 = \alpha_1, x_3 - x_4 = z(f_{x_2} - f_{x_1})/k, x_4 - x_5 = \alpha_2, \ldots, \\
&\quad x_{2n-3} - x_{2(n-1)} = z(f_{x_{n-1}} - f_{x_{n-2}})/k, \\
&\quad\qquad x_{2(n-1)} - x_{2n-1} = \alpha_{n-1}, \\
&\quad y_1 - y_2 = zf_{y_1}/k, y_2 - y_3 = \beta_1, y_3 - y_4 = z(f_{y_2} - f_{y_1})/k, \\
&\quad y_4 - y_5 = \beta_2, \ldots, y_{2n-3} - y_{2(n-1)} = z(f_{y_{n-1}} - f_{y_{n-2}})/k, \\
&\quad y_{2(n-1)} - y_{2n-1} = \beta_{n-1}] - 1\}. \qquad (27)
\end{aligned}$$

Equations (24) and (26) or (27) are the basic general formulas. We next consider the case of a wave which has weak random phase fluctuations across the initial plane. This "weak phase screen" case has previously been analyzed[3] directly from the $4n$-fold integral only in the case of Gaussian ρ.

2. WEAK PHASE SCREEN

Assuming unit amplitude and random phase, $S(\mathbf{r}_i) = S_i$, on the $z = 0$ plane

$$\begin{aligned}
M_0(\mathbf{r}_1, \ldots, \mathbf{r}_{2n}) &= \langle \psi_1\psi_2^*\psi_3\psi_4^* \cdots \psi_{2n-1}\psi_{2n}^*\rangle \\
&= \langle \exp j(S_1 - S_2 + S_3 - S_4 + \cdots + S_{2n-1} - S_{2n})\rangle \\
&= \exp - \tfrac{1}{2}\langle S_1^2 + S_2^2 + S_2^2 + \cdots + S_{2n}^2 \\
&\quad + \sum_{\substack{i=1\\ i\neq j}}^{2n}\sum_{j=1}^{2n} (-1)^{i+j} S_i S_j\rangle, \qquad (28)
\end{aligned}$$

where we assume Gaussian variables. We define the phase autocorrelation function $\rho_{ij} = \rho(r_i, r_j)$:

$$\langle S_i S_j\rangle = S_0^2 \rho_{ij}, \qquad (29)$$

where $\rho_{ii} = 1$ and, because we assume homogeneity,

$$\rho_{ij} = \rho([(x_i - x_j)^2 + (y_i - y_j)^2]^{1/2}). \qquad (30)$$

Thus

$$\begin{aligned}
&M_0(\mathbf{r}_1, \ldots, \mathbf{r}_{2n}) \\
&\quad = \exp\{- S_0^2[n - (\rho_{12} - \rho_{13} + \rho_{14} - \cdots + \rho_{1\,2n}) \\
&\quad - (\rho_{23} - \rho_{24} + \rho_{25} - \cdots - \rho_{2\,2n}) \\
&\quad - (\rho_{34} - \rho_{35} + \rho_{36} - \cdots + \rho_{3\,2n}) \\
&\quad - (\rho_{45} - \rho_{46} + \rho_{47} - \cdots - \rho_{4\,2n}) - \cdots \\
&\quad - (\rho_{2n-1\,2n})]\}. \qquad (31)
\end{aligned}$$

For a weak phase screen, $nS_0^2 \ll 1$, we write

$$\begin{aligned}
&M_0(\mathbf{r}_1, \ldots, \mathbf{r}_{2n}) \\
&\quad \approx 1 + S_0^2[- n + (\rho_{12} - \rho_{13} + \rho_{14} - \cdots + \rho_{1\,2n}) \\
&\quad + (\rho_{23} - \rho_{24} + \rho_{25} - \cdots - \rho_{2\,2n}) \\
&\quad + (\rho_{34} - \rho_{35} + \rho_{36} - \cdots - \rho_{32n}) + \cdots + (\rho_{2n-1\,2n})], \qquad (32)
\end{aligned}$$

and obtain for the nth moment of the intensity fluctuations, substituting in Eq. (24) and using Eq. (32),

$$\begin{aligned}
&\langle I^n(z)\rangle \\
&\quad = \left(\frac{k}{2\pi z}\right)^{2(n-1)} S_0^2 \int^{(4[n-1])} d\xi_1 d\xi_2 \cdots d\xi_{2(n-1)} d\xi_{2n} \\
&\quad\times d\xi_{2n+1} \cdots d\xi_{4n-3} \exp\left[\frac{jk}{z}\left\{\sum_{i=1}^{n-1}\sum_{m=0}^{i-1} \xi_{2i}\xi_{2m+1}\right.\right. \\
&\quad \left.\left. + \sum_{i=1}^{n-1}\sum_{m=0}^{i-1} \xi_{2(n+i)-1}\xi_{2(m+n)}\right\}\right] \{(S_0^{-2} - n) + (\rho_{12} - \rho_{13} \\
&\quad + \rho_{14} - \cdots + \rho_{1\,2n}) + (\rho_{23} - \rho \quad + \rho_{25} - \cdots - \rho_{2\,2n}) \\
&\quad + \cdots + (\rho_{2n-1\,2n})\}_R, \qquad (33)
\end{aligned}$$

where R signifies that this term is evaluated using the replacements

$$\left.\begin{aligned}&\xi_{2i} \to -\xi_{2i}\\ &\xi_{2(n+i)-1} \to -\xi_{2(n+i)-1}\end{aligned}\right\} \quad \text{for } i = 1, 2, \ldots, n-1;$$

$$p_i = q_i = 0 \quad \text{for } i = 1, 2, \ldots, 2(n-1).$$

Using our definitions of ξ, ρ_{ij} can be expressed as

$$\rho_{ij} = \rho\left(\left\{\left[\sum_{m=i}^{j-1} (-1)^{m+1}\xi_m\right]^2 + \left[\sum_{m=0}^{j-2} (-1)^m \xi_{2n+m}\right]^2\right\}^{1/2}\right). \tag{34}$$

Note that we must take

$$\xi_{2n-1} = -(\xi_1 + \xi_3 + \xi_5 + \ldots + \xi_{2n-3}),$$
$$\xi_{2(2n-1)} = -(\xi_{2n} + \xi_{2(n+1)} + \xi_{2(n+2)} + \ldots + \xi_{4(n-1)}). \tag{35}$$

Equation (33) consists of a sum of $n(2n-1)+1$ integrals resulting from the $n(2n-1)+1$ terms in curly brackets which are labelled $J_0, J_1, J_2, \ldots, J_{n(2n-1)}$, respectively. The integral J_0 resulting from the contribution of the term $(S_0^{-2} - n)$ can be evaluated by first integrating over the variables ξ_{2i} and $\xi_{2(n+i)-1}$ which appear in the double sums in the exponential. We find

$$\begin{aligned}J_0 = (1 - nS_0^2)(k/2\pi z)^{2(n-1)} &\int^{(4[n-1])} d\xi_1 d\xi_2 \cdots\\ &\times d\xi_{2(n-1)} d\xi_{2n}\, d\xi_{2n+1} \cdots d\xi_{4n-3}\\ &\times \exp\{(jk/z)[\xi_2\xi_1 + \xi_4(\xi_1 + \xi_3) + \xi_6(\xi_1 + \xi_3 + \xi_5)\\ &+ \ldots + \xi_{2(n-1)}(\xi_1 + \xi_3 + \xi_5 + \ldots + \xi_{2n-3})]\\ &+ [\xi_{2n+1}\xi_{2n} + \xi_{2n+3}(\xi_{2n} + \xi_{2(n+1)}) + \xi_{2n+3}(\xi_{2n}\\ &+ \xi_{2(n+1)} + \xi_{2(n+2)}) + \cdots\\ &+ \xi_{4n-3}(\xi_{2n} + \xi_{2(n+1)} + \xi_{2(n+2)} + \ldots + \xi_{4(n-1)})]\}.\end{aligned} \tag{36}$$

Integrating over $\xi_2, \xi_4, \ldots, \xi_{2(n-1)}, \xi_{2n+1}, \xi_{2n+3}, \ldots, \xi_{4n-3}$, we obtain a series of δ functions with the result

$$J_0 = 1 - nS_0^2. \tag{37}$$

The integral J_1, containing the term $\rho((\xi_1^2 + \xi_{2n}^2)^{1/2})$, can be evaluated by first integrating over $\xi_2, \xi_4, \ldots, \xi_{2(n-1)}$, $\xi_{2n+1}, \xi_{2n+3}, \ldots, \xi_{4n-3}$ to obtain

$$\begin{aligned}J_1 = S_0^2 &\int^{(2[n-1])} d\xi_1 d\xi_3 \cdots d\xi_{2n-3} d\xi_{2n} d\xi_{2(n+1)} \cdots\\ &\times d\xi_{4(n-1)} \rho((\xi_1^2 + \xi_{2n}^2)^{1/2}) \delta(\xi_1)\delta(\xi_1 + \xi_3) \cdots\\ &\times \delta(\xi_1 + \xi_3 + \cdots + \xi_{2n-3})\delta(\xi_{2n})\delta(\xi_{2n} + \xi_{2(n+1)})\\ &\times \cdots \delta(\xi_{2n} + \xi_{2(n+1)} + \cdots + \xi_{4(n-1)})\\ &= S_0^2.\end{aligned} \tag{38}$$

The integral J_2, which contains $\rho_{13} = \rho[(\xi_1 - \xi_2)^2 + (\xi_{2n} - \xi_{2n+1})^2]^{1/2}$, can be obtained by first integrating over $\xi_4, \xi_6, \ldots, \xi_{2(n-1)}, \xi_{2n+3}, \xi_{2n+5}, \ldots, \xi_{4n-3}$ to obtain

$$\begin{aligned}J_2 = &-(k/2\pi z)^2 S_0^2\\ &\times \int^{(2n)} d\xi_1 d\xi_2 d\xi_3 d\xi_5 \cdots d\xi_{2n-3} d\xi_{2n} d\xi_{2n+1}\\ &\times d\xi_{2(n+1)} d\xi_{2(n+2)} \cdots d\xi_{4(n-1)}\\ &\times \rho[(\xi_1 - \xi_2)^2 + (\xi_{2n} - \xi_{2n+1})^2]^{1/2}\\ &\times \exp[(jk/z)(\xi_2\xi_1 + \xi_{2n+1}\xi_{2n})]\\ &\times \delta(\xi_1 + \xi_3)\delta(\xi_1 + \xi_3 + \xi_5) \cdots\\ &\times \delta(\xi_1 + \xi_3 + \xi_5 + \cdots + \xi_{2n-3})\\ &\times \delta(\xi_{2n} + \xi_{2(n+1)})\delta(\xi_{2n} + \xi_{2(n+1)} + \xi_{2(n+2)})\\ &\times \delta(\xi_{2n} + \xi_{2(n+1)} + \xi_{2(n+2)} + \xi_{2(n+3)}) \cdots\\ &\times \delta(\xi_{2n} + \xi_{2(n+1)} + \cdots + \xi_{4(n-1)})\\ &= -(k/2\pi z)^2 S_0^2 \int^{(4)} d\xi_1 d\xi_2 d\xi_{2n} d\xi_{2n+1} \rho[(\xi_1 - \xi_2)^2\\ &+ (\xi_{2n} - \xi_{2n+1})^2]^{1/2}) \exp[(jk/z)(\xi_1\xi_2 + \xi_{2n}\xi_{2n+1})].\end{aligned} \tag{39}$$

By setting $A = \xi_1 - \xi_2$, $B = \xi_1 + \xi_2$, $C = \xi_{2n} - \xi_{2n+1}$, $D = \xi_{2n} + \xi_{2n+1}$, the integrations over B and D yield δ functions with the result

$$\begin{aligned}J_2 = -j(k/4\pi z) S_0^2 &\int^{(2)} dA dC \rho([A^2 + C^2]^{1/2})\\ &\times \exp[-(jk/4z)(A^2 + C^2)].\end{aligned} \tag{40}$$

Each of the remaining integrals in Eq. (33), $J_3, J_4, \ldots, J_{n(2n-1)}$, can similarly be reduced to one of the forms given above for J_1 or J_2:

(a) The n^2 integrals, including J_1, which contain ρ_{ij} with $i + j$ odd each similarly yield S_0^2.

(b) The $(n/2)(n-1)$ integrals, including J_2, which contain ρ_{ij} with i and j odd each similarly yield the expression for J_2 given above.

(c) The $(n/2)(n-1)$ integrals which contain ρ_{ij} with i and j even each similarly yield J_2^*.

Combining all the results, we obtain

$$\begin{aligned}\langle I^n(z)\rangle &= (1 - nS_0^2) + n^2 S_0^2 + (n/2)(n-1)(J_2 + J_2^*)\\ &= 1 + n(n-1)[S_0^2 + \mathrm{Re}(J_2)]\\ &= 1 + n(n-1)S_0^2[1 - \textstyle\int_0^\infty \rho(2\sqrt{zt/k})\sin t dt],\end{aligned} \tag{41}$$

where the expression for J_2 has been simplified by transforming to polar coordinates (r, φ) in the A–C plane, integrating over φ, and then setting $t = kr^2/4z$. For a Gaussian model for the phase autocorrelation, $\rho(r) = \exp(-r^2/l^2)$, we obtain

$$\langle I^n(z)\rangle = 1 + n(n-1)S_0^2 (4z/kl^2)^2/[1 + (4z/kl^2)^2], \tag{42}$$

which applies to all distances z for $nS_0^2 \ll 1$. [Mercier[4] also evaluated all moments of the intensity distribution with Gaussian autocorrelation of phase variation for the limit $z \to \infty$ and any S_0^2. The special result that appears in his paper and Eq. (40) reduce to $\langle I^n(z)\rangle = 1 + n(n-1)S_0^2$ for the common case of $z \to \infty$, Gaussian ρ, and $nS_0^2 \ll 1$.]

To obtain the nth-order intensity autocorrelation $\langle I(p_1, q_1, z) I(p_2, q_2, z) \cdots I(p_{n-1}, q_{n-1}, z) I(0, 0, z)\rangle$ in the transverse plane we need only recalculate Eq. (33), substituting $\xi_{2i} - p_i$ for ξ_{2i} and $\xi_{2(n+i)-1} - q_i$ for $\xi_{2(n+i)-1}$, where $i = 1, 2, 3, \ldots, n-1$, wherever they appear in the argument of ρ in the integral in Eq. (33). We obtain the following result upon carrying through this procedure:

$$\begin{aligned}&\langle I(p_1, q_1, z) \cdots I(p_{n-1}, q_{n-1}, z) I(0, 0, z)\rangle\\ &= 1 + 2S_0^2 \sum_{i=1}^{n-1}\left\{\sum_{m=i}^{n-1} \rho\left(\left[\left(\sum_{k=i}^{m} p_k\right)^2 + \left(\sum_{k=i}^{m} q_k\right)^2\right]^{1/2}\right)\right.\\ &\quad - \frac{k}{4\pi z}\sum_{m=i}^{n-1} \int^{(2)} dA dC \rho\left(\left[\left(A + \sum_{k=i}^{m} p_k\right)^2\right.\right.\\ &\quad \left.\left.\left. + \left(C + \sum_{k=i}^{m} q_k\right)^2\right]^{1/2}\right) \sin\left(\frac{k}{4z}(A^2 + C^2)\right)\right\}.\end{aligned} \tag{43}$$

For a Gaussian model, the integrals involved may be readily evaluated. Setting $\rho(x, y) = \exp(-x^2/l_x^2 - y^2/l_y^2)$, we obtain

$$\langle I(p_1,q_1,z)I(p_2,q_2,z)\cdots I(p_{n-1},q_{n-1},z)I(0,0,z)\rangle$$
$$= 1 + 2S_0^2 \sum_{i=1}^{n-1}\left(\sum_{m=i}^{n-1}\left\{\exp\left[-\left(\sum_{k=i}^{m} p_k\right)^2 l_x^{-2} - \left(\sum_{k=i}^{m} q_k\right)^2 l_y^{-2}\right]\right.\right.$$
$$- \left(\frac{Q_x^2\cdot Q_y^2}{(1+Q_x^2)(1+Q_y^2)}\right)^{1/4} \exp\left[-\frac{Q_x^2}{l_x^2(1+Q_x^2)}\left(\sum_{k=i}^{m} p_k\right)^2\right.$$
$$\left.- \frac{Q_y^2}{l_y^2(1+Q_y^2)}\left(\sum_{k=i}^{m} q_k\right)^2\right] \sin\left[\tfrac{1}{2}\tan^{-1}Q_x + \tfrac{1}{2}\tan^{-1}Q_y\right.$$
$$\left.\left. + \frac{Q_x}{l_x^2(1+Q_x^2)}\left(\sum_{k=i}^{m} p_k\right)^2 + \frac{Q_y}{l_y^2(1+Q_y^2)}\left(\sum_{k=i}^{m} q_k\right)^2\right]\right\}\Bigg), \tag{44}$$

where $Q_x = kl_x^2/4z$, $Q_y = kl_y^2/4z$.

The nth-order spectrum of intensity fluctuations may be obtained from Eq. (26). After some manipulation, we find

$$\hat{M}_{nI}(f_{x_1}, f_{x_2}, \ldots, f_{x_{n-1}}, f_{y_1}, f_{y_2}, \ldots, f_{y_{n-1}}, z)$$
$$= S_0^2\left(\sum_{i=1}^{n-1}\left\{\sum_{m=i}^{n-1}(2\pi)^{2(n-m-1)}\hat{\rho}\left(\sum_{j=1}^{m} f_{x_j}, \sum_{j=1}^{m} f_{y_j}\right)\right.\right.$$
$$\times\left[\exp\left(\frac{jz}{k}\right)[f_{x_i}(f_{x_m} - f_{x_i}) + f_{y_i}(f_{y_m} - f_{y_i})]\right.$$
$$- \exp\left(\frac{jz}{k}\right)[f_{x_i}(f_{x_m} - f_{x_{i-1}}) + f_{y_i}(f_{y_m} - f_{y_{i-1}})]$$
$$+ \exp\left(\frac{jz}{k}\right)[f_{x_i}(f_{x_{m+1}} - f_{x_{i-1}}) + f_{y_i}(f_{y_{m+1}} - f_{y_{i-1}})]$$
$$\left.\left. - \exp\left(\frac{jz}{k}\right)[f_{x_i}(f_{x_{m+1}} - f_{x_i}) + f_{y_i}(f_{y_{m+1}} - f_{y_i})]\right]\right\}\Bigg), \tag{45}$$

where $\hat{\rho}(f_{x_i}, f_{y_i})$ is the Fourier transform of $\rho(\alpha_i, \beta_i)$.

For $n = 2$, we readily obtain the important formula first stated by Salpeter,[5]

$$\hat{M}_{2I}(f_{x_1}, f_{y_1}, z) = 4S_0^2\,\hat{\rho}(f_{x_1}, f_{y_1})\sin^2[z(f_{x_1}^2 + f_{y_1}^2)/2k]. \tag{46}$$

3. STATISTICAL PROPERTIES OF THE INTENSITY FIELD IN THE WEAK PHASE SCREEN CASE

For the weak phase screen model we have chosen, $\langle I\rangle = 1$, and we can show from Eq. (41) that the skewness $E[(I - \langle I\rangle)^3]/\sigma^3 \equiv 0$. Thus the distribution of I is symmetrical as expected. We can also demonstrate that the variance, σ^2, is small. From Eq. (41),

$$\sigma^2 \equiv \langle I^2(z)\rangle - \langle I(z)\rangle^2 = 2S_0^2\left[1 - \int_0^\infty \rho(2\sqrt{zt/k})\,\sin t\,dt\right]. \tag{47}$$

Assuming that ρ is a monotonically decreasing function with $\rho(\infty) = 0$, the integral is positive. Consequently, because σ^2 is positive, the term in brackets in Eq. (47) must lie between zero and one. Hence $\sigma^2 \le 2S_0^2 \ll 1$ for all weak phase screens. Salpeter,[5] employing the spectrum of intensity fluctuations, has also shown that in the case of $kl^2/z \ll 1$ and $kl^2/z \gg 1$, the variance, σ^2, is much less than unity. (In the case of a strong phase screen, $S_0^2 \ge 1$, and Gaussian autocorrelation for ρ, Mercier[4] expanded the exponential form of the fourth-order mutual coherence function on the screen in power series up to S_0^{32} and programmed the eightfold integral for numerical calculations. The numerical results show peaks corresponding to $\sigma^2 > 1$.)

We can now show that the distribution function of the intensity fluctuations is Gaussian to a good approximation. The nth moment of a Gaussian distribution with unity mean and variance σ^2 is[12]

$$m_n(\sigma^2) = 1 + \frac{n(n-1)}{2}\int_0^{\sigma^2} m_{n-2}(\sigma^2)\,d(\sigma^2). \tag{48}$$

For a Gaussian distribution with $m_0 = m_1 = 1$ and $\sigma^2 \ll 1$, Eq. (48) reduces to

$$m_n(\sigma^2) = 1 + \tfrac{1}{2}n(n-1)\sigma^2 + O(\sigma^4) \approx 1 + \tfrac{1}{2}n(n-1)\sigma^2. \tag{49}$$

Our expression for $\langle I^n(z)\rangle$ is identical to Eq. (49). Remembering that we require $nS_0^2 \ll 1$, we find therefore that the distribution of intensity is Gaussian in the limit $S_0 \to 0$ for all z. Mercier[4] derived the same result for $z \to \infty$ in the special case of Gaussian ρ. The Gaussian statistics of intensity fluctuations have been described previously by other authors[5,10] for the weak phase screen case. This same result can also be obtained from the formalism of the method of smooth perturbations given in the work of Tatarski[13] as a limit of the log-normal distribution for weak scintillations.

[1] D. J. Torrieri and L. S. Taylor, J. Math. Phys. **12**, 1211 (1971).
[2] L. S. Taylor, J. Math. Phys. **13**, 590 (1972).
[3] M. J. Beran in *Statistical Methods in Radio Wave Propagation*, edited by N. C. Hoffman (Pergamon, New York, 1960).
[4] R. P. Mercier, Proc. Camb. Philos. Soc. **58**, 382 (1962).
[5] E. E. Salpeter, Astrophys. J. **147**, 433 (1967).
[6] J. R. Jokipii, Astrophys. J. **161**, 1147 (1970).
[7] V. I. Shishov, Izv. Vuz. Radiofiz. **14**, 85 (1971).
[8] K. S. Gochelashvily and V. I. Shishov, Opt. Acta **18**, 313 (1971).
[9] K. S. Gochelashvily and V. I. Shishov, P. N. Lebedev Physical Institute, preprint no. 171 (1971).
[10] A. T. Young, Astrophys. J. **168**, 543 (1971).
[11] J. R. Jokipii and L. C. Lee, Astrophys. J. **172**, 729 (1972).
[12] A. Papoulis, *Probability, Random Variables, and Stochastic Processes* (McGraw-Hill, New York, 1965).
[13] V. I. Tatarski, *Wave Propagation in a Turbulent Medium*, transl. by R. A. Silverman (Dover, New York, 1961), pp. 208–210.

Wigner and Racah coefficients for SU_3*

J. P. Draayer† and Yoshimi Akiyama‡

Department of Physics, The University of Michigan, Ann Arbor, Michigan 48105
(Received 19 March 1973; revised manuscript received 22 May 1973)

A general yet simple and hence practical algorithm for calculating $SU_3 \supset SU_2 \times U_1$ Wigner coefficients is formulated. The resolution of the outer multiplicity follows the prescription given by Biedenharn and Louck. It is shown that SU_3 Racah coefficients can be obtained as a solution to a set of simultaneous equations with unknown coefficients given as a by-product of the initial steps in the $SU_3 \supset SU_2 \times U_1$ Wigner coefficient construction algorithm. A general expression for evaluating $SU_3 \supset R_3$ Wigner coefficients as a sum over a simple subset of the corresponding $SU_3 \supset SU_2 \times U_1$ Wigner coefficients is also presented. State conjugation properties are discussed and symmetry relations for both the $SU_3 \supset SU_2 \times U_1$ and $SU_3 \supset R_3$ Wigner coefficients are given. Machine codes based on the results are available.

1. INTRODUCTION

The work of Wigner on the theory of group representations[1] coupled with Racah's development of the algebra of tensor operators[2] provides basic simplifying techniques for spectroscopic analyses. The usefulness of their techniques in any particular situation, however, depends to a great extent upon the availability of the appropriate Wigner and Racah coefficients. Ordinary angular momentum algebra, for example, owes its utility as a calculational tool to the ready availability of SU_2 Wigner and Racah coefficients. Other more complicated group structures for which Wigner and Racah coefficients are not so readily available, however, are also known to have real physical significance. The special unitary group in three dimensions, SU_3, is a case in point. In 1958 Elliott pointed out its usefulness in understanding the rotational structure of light nuclei.[3] Some four years later it was also recognized as being of importance in the classification of elementary particles.[4] As a consequence, Wigner and Racah coefficients for this group have been given in either algebraic or numeric form for simple cases of special interest by a number of authors.[5] More general results have only recently been made available through the work of Biedenharn and Louck and co-workers.[6-12] Except for the case of multiplicity free and the so-called Γ_s couplings, however, an additional algorithm is needed if numerical values for Wigner coefficients are to be extracted from the formalism. And since most authors disagree in their choice of a phase convention, extreme caution must be used if results so obtained are used to augment simple algebraic formulas currently available. An additional complication exists because two inequivalent reductions are needed: $SU_3 \supset SU_2 \times U_1$ in particle physics and $SU_3 \supset R_3$ in nuclear physics.

The purpose of this article is to: (i) Formulate in the spirit of an ordinary tensor formalism (built with tensors which by construction have the same null space properties as the Biedenharn and Louck Wigner operators) a general but simple and hence practical algorithm for generating $SU_3 \supset SU_2 \times U_1$ Wigner coefficients for arbitrary couplings and multiplicities; (ii) express SU_3 Racah coefficients as the solution to a set of simultaneous equations with the unknown coefficients given as a by-product of the initial steps in the $SU_3 \supset SU_2 \times U_1$ Wigner coefficient construction algorithm; (iii) exploit properties of the $SU_3 \supset R_3$ projection process together with known transformation coefficients between the $SU_3 \supset SU_2 \times U_1$ and $SU_3 \supset R_3$ schemes to express $SU_3 \supset R_3$ Wigner coefficients as a sum over a particularly simple subset of the corresponding $SU_3 \supset SU_2 \times U_1$ Wigner coefficients; (iv) list symmetry properties of the transformation coefficients between the $SU_3 \supset SU_2 \times U_1$ and $SU_3 \supset R_3$ schemes and discuss conjugation properties of state vectors for both reductions; (v) give symmetry properties for both the $SU_3 \supset SU_2 \times U_1$ and $SU_3 \supset R_3$ Wigner coefficients. We begin by briefly reviewing common notations and discussing their relationship to one another.

2. BASIC NOTATION

The labels λ and μ are used to characterize the irreducible representations of SU_3. The row labels in the $SU_3 \supset SU_2 \times U_1$ reduction are chosen as

$$\begin{aligned} \epsilon &= 2\lambda + \mu - 3(p+q) = -3Y, \\ \Lambda &= (\mu + p - q)/2 = I, \\ M_\Lambda &= r - \Lambda = I_z, \end{aligned} \tag{1}$$

where the integers p, q, r satisfy $0 \le p \le \lambda$, $0 \le q \le \mu$, $0 \le r \le 2\Lambda$. The notation $|(\lambda\mu)\epsilon\Lambda M_\Lambda\rangle$ is that introduced by Elliott into nuclear physics to label states in the so-called intrinsic or body-fixed system.[3,13] In terms of a three-dimensional oscillator with n_i quanta in the i-direction, $\epsilon = 2n_3 - n_1 - n_2$ while Λ labels the irreducible representation of SU_2 with projection $M_\Lambda = (n_1 - n_2)/2$. In particle physics states are labeled as $|(\lambda\mu)YII_z\rangle$ with Y denoting the hypercharge and I and I_z the isospin and its projection, respectively.[14]

An equivalent but mathematically more elegant notation is that due to Gel'fand in which case states are labeled by patterns of the type[15]

$$|G\rangle \equiv \left| \begin{matrix} g_{13} & g_{23} & g_{33} \\ & g_{12} & g_{22} \\ & & g_{11} \end{matrix} \right\rangle . \tag{2}$$

The g_{ij}, $1 \le i \le j \le 3$, specify the irreducible representation of U_j in the chain $U_3 \supset U_2 \supset U_1$. Specifically, g_{ij} is the number of boxes in row i of the Young tableau for U_j. $\lambda = g_{13} - g_{23}$, $\mu = g_{23} - g_{33}$, and $\nu = g_{33}$ are then the number of columns containing 1, 2, and 3 boxes, respectively, in the Young tableau for U_3. For notational convenience G (for Gel'fand) will be used to denote the full set of g_{ij} labels. Apart from an n_i-dependent phase factor $|G\rangle = |(\lambda\mu)\epsilon\Lambda M_\Lambda\rangle$ with $g_{12} = p + \mu + \nu = \frac{1}{3}(\lambda + 2\mu) - \frac{1}{6}\epsilon + \Lambda + \nu$, $g_{22} = q + \nu = \frac{1}{3}(\lambda + 2\mu) - \frac{1}{6}\epsilon - \Lambda + \nu$, $g_{11} = r + q + \nu = 2M_\Lambda + \frac{1}{3}(\lambda + 2\mu) - \frac{1}{6}\epsilon - 3\Lambda + \nu$. The so-called betweenness conditions ($g_{ij} \ge g_{i,j-1} \ge g_{i+1,j}$) are equivalent to the restrictions $0 \le p \le \lambda$, $0 \le q \le \mu$, $0 \le r \le 2\Lambda$.

States of particular interest are those for which the number of oscillator quanta ($n_i = \sum_j g_{j,i} - \sum_j g_{j,i-1}$) in

TABLE I. (Subgroup labels for extremal states). The subscripts HW and LW mean highest weight and lowest weight in the Gel'fand sense (not to be confused with $\epsilon_{max} \sim G_{LW}$ and $\epsilon_{min} \sim G_{HW}$).

G	g_{12}	g_{22}	g_{11}	p	q	r	ϵ	2Λ	$2M_\Lambda$	n_1	n_2	n_3	I	J
G_{HW}	g_{13}	g_{23}	g_{13}	λ	μ	λ	$-\lambda-2\mu$	λ	λ	$\lambda+\mu+\nu$	$\mu+\nu$	ν	1	1
G_{HW}	g_{13}	g_{23}	g_{23}	λ	μ	0	$-\lambda-2\mu$	λ	$-\lambda$	$\mu+\nu$	$\lambda+\mu+\nu$	ν	1	0
G_{LW}	g_{23}	g_{33}	g_{33}	0	0	0	$2\lambda+\mu$	μ	$-\mu$	ν	$\mu+\nu$	$\lambda+\mu+\nu$	0	0
G_{LW}	g_{23}	g_{33}	g_{23}	0	0	μ	$2\lambda+\mu$	μ	μ	$\mu+\nu$	ν	$\lambda+\mu+\nu$	0	1

the 3-direction is either a maximum or a minimum. The value of the subgroup labels for these so-called extremal states ($|G_E\rangle$) are summarized by Table I. The I and J labels form a convenient code by which the states can be distinguished. The labels λ, μ, ν can therefore be thought of as either specifying or being specified by the distribution of oscillator quanta for extremal states.

In the $SU_3 \supset R_3$ reduction states are labeled by the total angular momentum L and its projection M. Multiple occurrences of a given L can be distinguished in a variety of ways.[16] The physically most significant scheme is that due to Elliott in which case K, the projection of L along the body-fixed 3-axis, is used to sort the L-values into the familiar K-bands of rotational model theory.[17] The prescription given is that projected states defined by

$$|(G)KLM\rangle \equiv P^L_{MK}|G\rangle \equiv (2L+1)\int d\Omega\ D^{L*}_{MK}(\Omega)R(\Omega)|G\rangle \quad (3)$$

form a complete basis if $G = G_E$ and for:

$$G_E = G_{HW}: \quad K = \lambda, \lambda-2, \ldots, 1 \text{ or } 0,\quad L = K, K+1, \ldots, K+\mu, \quad K \neq 0,\quad L = \mu, \mu-2, \ldots, 1 \text{ or } 0, \quad K = 0; \quad (4a)$$

$$G_E = G_{LW}: \quad K = \mu, \mu-2, \ldots, 1 \text{ or } 0,\quad L = K, K+1, \ldots, K+\lambda, \quad K \neq 0,\quad L = \lambda, \lambda-2, \ldots, 1 \text{ or } 0, \quad K = 0. \quad (4b)$$

In Eq. (3), $D^L_{MK}(\Omega)$ is an R_3 rotation matrix and $R(\Omega)$ is an R_3 rotation operator. The integration is over Euler angles.

States defined by Eqs. (3)–(4) are not normalized nor are they orthogonal with respect to the K-label. Working within such a scheme leads ultimately to nonhermitian matrices. To avoid this complication, it is convenient to orthonormalize the basis using a Gram–Schmidt process. The physical interpretation of K as a band label can be maintained approximately if a prescription analogous to that outlined by Vergados is used.[18] In this case

$$|(G_E)\mathcal{K}_i LM\rangle = \sum_{j\le i} O_{ij}|(G_E)K_j LM\rangle, \quad (5)$$

where the orthonormalization matrix O_{ij} is defined recursively by the formulas

$$O_{ii} = 1/(\langle(G_E)K_i LM|(G_E)K_i LM\rangle - \sum_{j<i} O_{ji}O_{ji})^{1/2} \quad (6a)$$

$$O_{ji} = O_{jj}(\langle(G_E)K_j LM|(G_E)K_i LM\rangle - \sum_{k<j<i} O_{kj}O_{ki})^{1/2} \quad (6b)$$

$$O_{ij} = O_{ii}(\delta_{ij} - \sum_{j\le k<i} O_{kj}O_{ki}). \quad (6c)$$

An analytic expression which allows the coefficients $\langle(G_E)K_i LM|(G_E)K_j LM\rangle$ to be evaluated is given in Sec. 3. Unlike the κ of Vergados, $\mathcal{K}$ like K is given by either Eq. (4a) or Eq. (4b). The extent to which different K-values are mixed by the orthonormalization process depends upon the relative magnitude of the coefficients $\langle(G_E)K_i LM|(G_E)K_i LM\rangle$ and $\langle(G_E)K_i LM|(G_E)K_j LM\rangle$. It can be verified that the mixing is indeed small. In particular, for $G_E = G_{HW}(G_{LW})$ and $i \neq j$ $\langle(G_E)K_i LM|(G_E)K_j LM\rangle \to 0$ if $\lambda(\mu)$ is fixed and $\mu(\lambda) \to \infty$.

3. ALGEBRAIC FORMULATION

If α represents a set of row labels used to distinguish orthonormal basis states within a given representation of $SU_3(\alpha = \epsilon\Lambda M_\Lambda$, or $\mathcal{K}LM$, or $\cdots)$, the Wigner coefficients $\langle(\lambda_1\mu_1)\alpha_1; (\lambda_2\mu_2)\alpha_2|(\lambda_3\mu_3)\alpha_3\rangle_\rho$ are by definition the elements of a unitary transformation between coupled and uncoupled representations of SU_3 in the α-scheme,

$$|(\lambda_3\mu_3)\alpha_3\rangle_\rho = \sum_{\alpha_1\alpha_2} \langle(\lambda_1\mu_1)\alpha_1; (\lambda_2\mu_2)\alpha_2|(\lambda_3\mu_3)\alpha_3\rangle_\rho |(\lambda_1\mu_1)\alpha_1\rangle|(\lambda_2\mu_2)\alpha_2\rangle. \quad (7)$$

The outer multiplicity label $\rho = 1, 2, \ldots, \rho_{max}$ is used to distinguish multiple occurrences of a given $(\lambda_3\mu_3)$ in the direct product $(\lambda_1\mu_1) \times (\lambda_2\mu_2)$. Although a definition bearing physical significance comparable, for example, to that associated with Elliott's choice of K for a resolution of the inner multiplicity problem in the $SU_3 \supset R_3$ reduction has not been proposed to fix ρ, Biedenharn and Louck and co-workers have demonstrated in a series of articles[6–12] that a mathematically canonical definition which puts the outer multiplicity on a sound group theoretical basis can be obtained through the use of the labels of an upper Gel'fand pattern for a Wigner operator of irreducible tensor character $(\lambda_2\mu_2)$. The practical aspects of this choice are manifest in the vanishing of certain Wigner and Racah coefficients [Eqs. (15), (23), below], simple symmetry relations under conjugation [Eqs. (32)–(36), below], and nice limiting properties for the $SU_3 \supset SU_2 \times U_1$ Wigner coefficients (see Ref. 11, for example). Outlined below are techniques which exploit the essential features of this definition (albeit somewhat obscured but only so as to minimize notational needs) in defining an algorithm (based on an ordinary tensor formalism built with tensor operators which by construction have the same null space properties as the Wigner operators of Biedenharn and Louck) which can be used to evaluate all $SU_3 \supset SU_2 \times U_1$ Wigner coefficients. Note that for most practical purposes, however, the outer multiplicity can be considered fully labelled with a running index $\rho = 1, 2, \ldots, \rho_{max}$ which distinguishes orthonormal basis states in the product space,

$$\sum_{\alpha_1\alpha_2} \langle(\lambda_1\mu_1)\alpha_1; (\lambda_2\mu_2)\alpha_2|(\lambda_3\mu_3)\alpha_3\rangle_\rho \times \langle(\lambda_1\mu_1)\alpha_1; (\lambda_2\mu_2)\alpha_2|(\lambda'_3\mu'_3)\alpha_3\rangle_{\rho'} = \delta_{\lambda'_3\lambda_3}\delta_{\mu'_3\mu_3}\delta_{\rho'\rho}. \quad (8)$$

A. $SU_3 \supset SU_2 \times U_1$ Wigner coefficients

Irreducible tensor operators under SU_3, $T^{(\lambda\mu)}$, can be defined through their commutation properties with the infinitesimal generators of the group.[19] The Wigner–Eckart theorem allows one to express the matrix elements of tensor operators defined in this manner as a sum over ρ of the product of a ρ-dependent generalized

reduced matrix element multiplied by the corresponding Wigner coefficient. Specifically, for the $SU_3 \supset SU_2 \times U_1$ reduction,

$$\langle(\lambda_3\mu_3)\epsilon_3\Lambda_3 M_{\Lambda_3}|\, T^{(\lambda_2\mu_2)}_{\epsilon_2\Lambda_2 M_{\Lambda_2}}\,|(\lambda_1\mu_1)\epsilon_1\Lambda_1 M_{\Lambda_1}\rangle$$
$$= \sum_\rho \langle\!\langle(\lambda_3\mu_3)\|\, T^{(\lambda_2\mu_2)}\,\|(\lambda_1\mu_1)\rangle\!\rangle_\rho \tag{9}$$
$$\times \langle(\lambda_1\mu_1)\epsilon_1\Lambda_1 M_{\Lambda_1};(\lambda_2\mu_2)\epsilon_2\Lambda_2 M_{\Lambda_2}|\,(\lambda_3\mu_3)\epsilon_3\Lambda_3 M_{\Lambda_3}\rangle_\rho.$$

This result can be used to define Wigner coefficients through the matrix elements of specially chosen tensor operators $K^{(\lambda_2\mu_2)}(\rho)$,

$$\langle(\lambda_3\mu_3)\epsilon_3\Lambda_3 M_{\Lambda_3}|K^{(\lambda_2\mu_2)}_{\epsilon_2\Lambda_2 M_{\Lambda_2}}(\rho)|\,(\lambda_1\mu_1)\epsilon_1\Lambda_1 M_{\Lambda_1}\rangle$$
$$= \langle(\lambda_3\mu_3)\|\, K^{(\lambda_2\mu_2)}(\rho)\|(\lambda_1\mu_1)\rangle \tag{10}$$
$$\times \langle(\lambda_1\mu_1)\epsilon_1\Lambda_1 M_{\Lambda_1};(\lambda_2\mu_2)\epsilon_2\Lambda_2 M_{\Lambda_2}|\,(\lambda_3\mu_3)\epsilon_3\Lambda_3 M_{\Lambda_3}\rangle_\rho$$

for which the ρ-summation of Eq. (9) is *not* required. The generalized reduced matrix element $\langle(\lambda_1\mu_1)\|K^{(\lambda_2\mu_2)}(\rho)\|(\lambda_3\mu_3)\rangle$ is then just a normalization factor. In particular, the infinitesimal generators which have irreducible tensor character $(\lambda_2\mu_2) = (11)$ and operate only within a given representation of SU_3 [e.g., $(\lambda_1\mu_1) = (\lambda_3\mu_3) = (\lambda\mu)$, only] are by definition matrix elements of the $\rho = 1$ variety.

The problem is then one of constructing the operators $K^{(\lambda_2\mu_2)}(\rho)$; and in particular, constructing them in a manner which serves to uniquely define the outer multiplicity label ρ. The scheme is straightforward: Clearly $\rho_{\max}$, the number of occurrences of $(\lambda_3\mu_3)$ in the direct product $(\lambda_1\mu_1) \times (\lambda_2\mu_2)$, depends upon $\lambda_1, \mu_1, \lambda_2, \mu_2, \lambda_3, \mu_3$. It is also clear that there exists an η such that $(\lambda_3\mu_3)$ occurs exactly ρ times in the product $(\lambda_1\mu_1) \times (\lambda_2 - \eta, \mu_2 - \eta)$. And in this case ρ depends upon $\lambda_1, \mu_1, \lambda_2 - \eta, \mu_2 - \eta, \lambda_3, \mu_3$. Let $\eta_{\max}$ be the value of η such that $(\lambda_1\mu_1) \times (\lambda_2 - \eta_{\max}, \mu_2 - \eta_{\max}) \to (\lambda_3\mu_3)$ is not allowed whereas $(\lambda_1\mu_1) \times (\lambda_2 - \eta_{\max} + 1, \mu_2 - \eta_{\max} + 1) \to (\lambda_3\mu_3)$ occurs with a multiplicity of one. Then $(\lambda_1\mu_1) \times (\bar\lambda_2 \equiv \lambda_2 - \eta_{\max} + \rho, \bar\mu_2 \equiv \mu_2 - \eta_{\max} + \rho) \to (\lambda_3\mu_3)$ occurs with a multiplicity of ρ for $\rho = 1, 2, \ldots, \rho_{\max} \le \eta_{\max}$. In this way, $(\lambda_1\mu_1) \times (\bar\lambda_2\bar\mu_2) \to (\lambda_3\mu_3)$ can be considered the parent coupling for the ρth occurrence of $(\lambda_3\mu_3)$ in the product $(\lambda_1\mu_1) \times (\lambda_2\mu_2)$. The question then arises: Is it possible to construct the $K^{(\lambda_2\mu_2)}(\rho)$ from the corresponding $K^{(\bar\lambda_2\bar\mu_2)}(\rho)$ in such a way as to preserve the unique null space) property of the parent operator which allow it to generate the ρth occurrence (and no more) of $(\lambda_3\mu_3)$ in the product space? The answer is yes, it can be done via a build-up process using the group generators $K^{(11)} \equiv K^{(11)}(\rho = 1)$. In particular, iterating the result

$$K^{(\lambda_2\mu_2)}_{\epsilon_2\Lambda_2 M_{\Lambda_2}}(\tilde\rho) = [K^{(\lambda_2-1,\mu_2-1)}(\rho) \times K^{(11)}]^{(\lambda_2\mu_2)}_{\epsilon_2\Lambda_2 M_{\Lambda_2}}$$
$$= \sum_{\epsilon\Lambda\Lambda'_2 M_\Lambda} \langle(11)\epsilon\Lambda M_\Lambda;(\lambda_2-1,\mu_2-1)$$
$$\epsilon'_2\Lambda'_2 M'_{\Lambda_2}\,|\,(\lambda_2\mu_2)\epsilon_2\Lambda_2 M_{\Lambda_2}\rangle\, K^{(\lambda_2-1,\mu_2-1)}_{\epsilon'_2\Lambda'_2 M'_{\Lambda_2}}(\rho) K^{(11)}_{\epsilon\Lambda M_\Lambda} \tag{11}$$

allows one to relate $K^{(\lambda_2\mu_2)}(\rho)$ to $K^{(\bar\lambda_2\bar\mu_2)}(\rho)$ for each ρ. Logical consistency demands, of course, that in each step $\tilde\rho$ be chosen numerically equal to ρ and that $\rho = 1$ corresponds to a multiplicity free parent coupling, $\rho = 2$ to the second solution in the parent coupling having a twofold outer multiplicity, etc. The tilde, however, is used to denote the fact that ρ-orthogonality in the product space is not guaranteed; that is, $K^{(\lambda_2\mu_2)}(\tilde\rho)$ will in general be a linear combination of all $K^{(\lambda_2\mu_2)}(\rho)$ with $\rho \le \tilde\rho$. (That operators with $\rho > \tilde\rho$ are not generated is a consequence of the fact that the group generators preserve the null space properties of the parent operator. Further discussion on the consequence of this result is given below. In effect, it means that the weight diagram[11] for a coupled operator of the type $T^{(\lambda\mu)} \times K^{(11)}$ is the same as for $T^{(\lambda\mu)}$.) To be sure, the build-up process cannot be used to define $K^{(\bar\lambda_2\bar\mu_2)}(\rho)$ because $K^{(\bar\lambda_2-1,\bar\mu_2-1)}(\rho) \equiv 0$. But this presents no major problem since an analytic expression for the Wigner coefficients corresponding to the ρth occurrence of $(\lambda_3\mu_3)$ in the product $(\lambda_1\mu_1) \times (\bar\lambda_2\bar\mu_2)$ is available [Eq. (20), below] and through Eq. (10) serves to define the first nonvanishing operator in the build-up process. Note that the Wigner coefficient appearing in Eq. (11) is multiplicity free. Substitution of Eq. (11) into Eq. (10) yields

$$\langle(\lambda_1\mu_1)\epsilon_1\Lambda_1 M_{\Lambda_1};(\lambda_2\mu_2)\epsilon_2\Lambda_2 M_{\Lambda_2}|\,(\lambda_3\mu_3)\epsilon_3\Lambda_3 M_{\Lambda_3}\rangle_{\tilde\rho}$$
$$= \langle(\lambda_3\mu_3)\|\, K^{(\lambda_2\mu_2)}(\tilde\rho)\|\,(\lambda_1\mu_1)\rangle^{-1}$$
$$\times \langle(\lambda_3\mu_3)\|K^{(\lambda_2-1,\mu_2-1)}(\rho)\,\|\,(\lambda_1\mu_1)\rangle\langle(\lambda_1\mu_1)\,\|\,K^{(11)}\,\|(\lambda_1\mu_1)\rangle$$
$$\times \sum_{\epsilon\Lambda\Lambda'_1\Lambda'_2 M_\Lambda} \langle(11)\epsilon\Lambda M_\Lambda;(\lambda_2-1,\mu_2-1)\epsilon'_2\Lambda'_2 M'_{\Lambda_2}|$$
$$(\lambda_2\mu_2)\epsilon_2\Lambda_2 M_{\Lambda_2}\rangle$$
$$\times \langle\!\langle(\lambda_1\mu_1)\epsilon_1\Lambda_1 M_{\Lambda_1};(11)\epsilon\Lambda M_\Lambda|\,(\lambda_1\mu_1)\epsilon'_1\Lambda'_1 M'_{\Lambda_1}\rangle\!\rangle_{\rho=1}$$
$$\times \langle(\lambda_1\mu_1)\epsilon'_1\Lambda'_1 M'_{\Lambda_1};(\lambda_2-1,\mu_2-1)\epsilon'_2\Lambda'_2 M'_{\Lambda_2}|$$
$$(\lambda_3\mu_3)\epsilon_3\Lambda_3 M_{\Lambda_3}\rangle_\rho. \tag{12}$$

If $K^{(11)}$ were not chosen to be of the generator type, representations other than $(\lambda_1\mu_1)$ would appear on the right- and left-hand sides of the matrix elements of $K^{(\lambda_2-1,\mu_2-1)}(\rho)$ and $K^{(11)}$ and a summation over these representation labels would be required. Factoring each coupling coefficient into a reduced coefficient (double-barred or isoscalar part) multiplied by an ordinary coefficient which carries the dependence upon the SU_2 projection labels and carrying out the summation over projection quantum numbers yields

$$\langle(\lambda_1\mu_1)\epsilon_1\Lambda_1;(\lambda_2\mu_2)\epsilon_2\Lambda_2\|\,(\lambda_3\mu_3)\epsilon_3\Lambda_3\rangle_{\tilde\rho}$$
$$= \langle(\lambda_3\mu_3)\|K^{(\lambda_2\mu_2)}(\tilde\rho)\,\|\,(\lambda_1\mu_1)\rangle^{-1}$$
$$\times \langle(\lambda_3\mu_3)\|K^{(\lambda_2-1,\mu_2-1)}(\rho)\,\|\,(\lambda_1\mu_1)\rangle\langle(\lambda_1\mu_1)\,\|\,K^{(11)}\|(\lambda_1\mu_1)\rangle$$
$$\times \sum_{\epsilon\Lambda\Lambda'_1\Lambda'_2} \langle(11)\epsilon\Lambda;(\lambda_2-1,\mu_2-1)\epsilon'_2\Lambda'_2\,\|\,(\lambda_2\mu_2)\epsilon_2\Lambda_2\rangle$$
$$\times \langle(\lambda_1\mu_1)\epsilon_1\Lambda_1;(11)\epsilon\Lambda\,\|\,(\lambda_1\mu_1)\epsilon'_1\Lambda'_1\rangle_{\rho=1}$$
$$\times \langle(\lambda_1\mu_1)\epsilon'_1\Lambda'_1;(\lambda_2-1,\mu_2-1)\epsilon'_2\Lambda'_2\|\,(\lambda_3\mu_3)\epsilon_3\Lambda_3\rangle_\rho$$
$$\times U(\Lambda_1\Lambda\Lambda_3\Lambda'_2;\Lambda'_1\Lambda_2), \tag{13}$$

where $U(\Lambda_1\Lambda\Lambda_3\Lambda'_2;\Lambda_1\Lambda_2)$ is an ordinary SU_2 recoupling coefficient and $\epsilon_1 = \epsilon_3 - \epsilon_2$, $\epsilon'_2 = \epsilon_2 - \epsilon$, $\epsilon'_1 = \epsilon_3 - \epsilon_2 + \epsilon$.

It should be emphasized that Eq. (13) is valid for completely general arguments $\epsilon_1, \Lambda_1, \epsilon_2, \Lambda_2, \epsilon_3, \Lambda_3$ and, furthermore, that certain coupling coefficients derived using this expression must necessarily vanish identically. To see this, consider in more detail a coefficient calculated by repeating the recursion process η times. The required matrix elements are for a tensor operator

$$K^{(\lambda_2\mu_2)}(\tilde{\rho}) = [\cdots[[K^{(\lambda_2-\eta,\mu_2-\eta)}(\rho)\times K^{(11)}]\times K^{(11)}]\cdots\times K^{(11)}]^{(\lambda_2\mu_2)}, \quad (14)$$

in which $K^{(11)}$ appears η times. In general the maximum change in Λ induced by an operator $K^{(\lambda\mu)}$ is $\frac{1}{2}(\lambda+\mu)$ since this is the maximum value of Λ in the representation $(\lambda\mu)$. The generators, however, are of a special type; they change Λ by at most $\frac{1}{2}$. The operator given by Eq. (14) can therefore change Λ by at most $(\lambda_2-\eta+\mu_2-\eta)/2+\eta/2=(\lambda_2+\mu_2-\eta)/2$. Consequently, the corresponding coupling coefficient must be zero if $|\Lambda_1-\Lambda_3|>\frac{1}{2}(\lambda_2+\mu_2-\eta)$. The maximum η for which this result is valid is simply $\eta=\eta_{\max}-\rho$. Consequently $\langle(\lambda_1\mu_1)\epsilon_1\Lambda_1;(\lambda_2\mu_2)\epsilon_2\Lambda_2\|(\lambda_3\mu_3)\epsilon_3\Lambda_3\rangle_{\tilde{\rho}}$ must vanish for $|\Lambda_1-\Lambda_3|>\frac{1}{2}(\lambda_2+\mu_2-\eta_{\max}+\tilde{\rho})$. This property is completely general and a direct consequence of the build-up process used to define the coefficients. Note that the number of coefficients predicted to be zero (more zeros may appear but for other reasons) is always a *decreasing* function of $\tilde{\rho}$. Although solutions obtained via repeated applications of Eq. (13) are not necessarily orthogonal with respect to the $\tilde{\rho}$-label, orthogonalizing in the *increasing* order $\rho=1,2,\ldots,\rho_{\max}$ using a Gram–Schmidt process preserves the vanishings; and hence the Wigner coefficients satisfy

$$\langle(\lambda_1\mu_1)\epsilon_1\Lambda_1;(\lambda_2\mu_2)\epsilon_2\Lambda_2\|(\lambda_3\mu_3)\epsilon_3\Lambda_3\rangle_\rho = 0 \quad \text{for } |\Lambda_1-\Lambda_3|>\tfrac{1}{2}(\lambda_2+\mu_2-\eta_{\max}+\rho). \quad (15)$$

This then quarantees the uniqueness of our result which by construction coincides with the Biedenharn and Louck prescription for a resolution of the outer multiplicity. Thus Eq. (13) provides a recursive means of defining the $SU_3\supset SU_2\times U_1$ Wigner coefficients for each mode of coupling characterized by ρ.

An expression which is computationally convenient to evaluate can be obtained from Eq. (13) by restricting $\epsilon_2\Lambda_2=\mathrm{HW}$ and $\epsilon_3\Lambda_3=\mathrm{HW}$. In this case $\epsilon\Lambda$ and $\epsilon_2'\Lambda_2'$ are also forced to be of HW and $\langle(11)\mathrm{HW};(\lambda_2-1,\mu_2-1)\mathrm{HW}\|(\lambda_2\mu_2)\mathrm{HW}\rangle=1$. The sum in Eq. (13) then reduces to simply

$$\sum_{\Lambda_1'}\langle(\lambda_1\mu_1)\epsilon_1\Lambda_1;(11)\mathrm{HW}\|(\lambda_1\mu_1)\epsilon_1-3,\Lambda_1'\rangle_{\rho=1}\times\langle(\lambda_1\mu_1)\epsilon_1-3,\Lambda_1';(\lambda_2-1,\mu_2-1)\mathrm{HW}\|(\lambda_3\mu_3)\mathrm{HW}\rangle_\rho\times U(\Lambda_1,\tfrac{1}{2},\lambda_3/2,(\lambda_2-1)/2;\Lambda_1',\lambda_2/2). \quad (16)$$

It follows (making use of results available, for example, in Refs. 19 and 20) that

$$\langle(\lambda_1\mu_1)\epsilon_1\Lambda_1;(\lambda_2\mu_2)\mathrm{HW}\|(\lambda_3\mu_3)\mathrm{HW}\rangle_{\tilde{\rho}} = N\left[-\left(\frac{R(p_1+1)A(\Lambda_1,\lambda_2/2,\lambda_3/2)[B(\Lambda_1,\lambda_2/2,\lambda_3/2)+1]}{(2\Lambda_1+2)(2\Lambda_1+1)}\right)^{1/2}\langle(\lambda_1\mu_1)\epsilon_1-3,\Lambda_1+\tfrac{1}{2};(\lambda_2-1,\mu_2-1)\mathrm{HW}\|(\lambda_3\mu_3)\mathrm{HW}\rangle_\rho + \left(\frac{S(q_1+1)C(\Lambda_1,\lambda_2/2,\lambda_3/2)D(\Lambda_1,\lambda_2/2,\lambda_3/2)}{(2\Lambda_1+1)(2\Lambda_1)}\right)^{1/2}\langle(\lambda_1\mu_1)\epsilon_1-3,\Lambda_1-\tfrac{1}{2};(\lambda_2-1,\mu_2-1)\mathrm{HW}\|(\lambda_3\mu_3)\mathrm{HW}\rangle_\rho\right],$$

$$\begin{aligned}
A(\Lambda_1,\Lambda_2,\Lambda_3)&=\Lambda_2+\Lambda_3-\Lambda_1,\\
B(\Lambda_1,\Lambda_2,\Lambda_3)&=\Lambda_3+\Lambda_1-\Lambda_2,\\
C(\Lambda_1,\Lambda_2,\Lambda_3)&=\Lambda_1+\Lambda_2-\Lambda_3,\\
D(\Lambda_1,\Lambda_2,\Lambda_3)&=\Lambda_1+\Lambda_2+\Lambda_3+1,\\
R(p_i)&=p_i(\lambda_i+1-p_i)(\mu_i+1+p_i),\\
S(q_i)&=q_i(\mu_i+1-q_i)(\lambda_i+\mu_i+2-q_i),
\end{aligned} \quad (17)$$

where N is a normalization factor. This result allows the recursion process of Eq. (13) to be carried out within a very limited number of coefficients. The restriction $\epsilon_3\Lambda_3=\mathrm{HW}$, however, also demands that

$$\langle(\lambda_1\mu_1)\epsilon_1\Lambda_1;(\lambda_2\mu_2)\epsilon_2+3,\Lambda_2'\|(\lambda_3\mu_3)\mathrm{HW}\rangle = \left(\frac{(2\Lambda_2'+1)}{(2\Lambda_1+1)(2\Lambda_2+1)N(\Lambda_2')}\right)^{1/2}\sum_{\Lambda_1'=\Lambda_1\pm1/2}\frac{X(\Lambda_1',\Lambda_2')}{\sqrt{2\Lambda_1'+1}}\times\langle(\lambda_1\mu_1)\epsilon_1+3,\Lambda_1';(\lambda_2\mu_2)\epsilon_2\Lambda_2\|(\lambda_3\mu_3)\mathrm{HW}\rangle,$$

$$\begin{aligned}
X(\Lambda_1+\tfrac{1}{2},\Lambda_2+\tfrac{1}{2})&=-\{S(q_1)[A(\Lambda_1,\Lambda_2,\lambda_3/2)+\tfrac{1}{2}]\times[B(\Lambda_1,\Lambda_2,\lambda_3/2)+\tfrac{1}{2}]\}^{1/2},\\
X(\Lambda_1-\tfrac{1}{2},\Lambda_2+\tfrac{1}{2})&=-\{R(p_1)[C(\Lambda_1,\Lambda_2,\lambda_3/2)+\tfrac{1}{2}]\times[D(\Lambda_1,\Lambda_2,\lambda_3/2)+\tfrac{1}{2}]\}^{1/2},\\
X(\Lambda_1-\tfrac{1}{2},\Lambda_2-\tfrac{1}{2})&=+\{S(q_1)[C(\Lambda_1,\Lambda_2,\lambda_3/2)+\tfrac{1}{2}]\times[D(\Lambda_1,\Lambda_2,\lambda_3/2)+\tfrac{1}{2}]\}^{1/2},\\
X(\Lambda_1-\tfrac{1}{2},\Lambda_2-\tfrac{1}{2})&=-\{R(p_1)[A(\Lambda_1,\Lambda_2,\lambda_3/2)+\tfrac{1}{2}]\times[B(\Lambda_1,\Lambda_2,\lambda_3/2)+\tfrac{1}{2}]\}^{1/2},
\end{aligned}$$

$$\begin{aligned}
N(\Lambda_2+\tfrac{1}{2})&=S(q_2),\\
N(\Lambda_2-\tfrac{1}{2})&=R(p_2).
\end{aligned} \quad (18)$$

And knowing this additional result allows all coefficients of the type $\epsilon_3\Lambda_3=\mathrm{HW}$ to be determined. Coefficients with $\epsilon_3\Lambda_3\neq\mathrm{HW}$ follow from the ordinary recursion formula

$$\begin{aligned}
\langle(\lambda_1\mu_1)\epsilon_1\Lambda_1;(\lambda_2\mu_2)\epsilon_2\Lambda_2\|(\lambda_3\mu_3)\epsilon_3\Lambda_3\rangle &= \frac{1}{N_3}\Bigg(\sum_{\Lambda_1'=\Lambda_1\pm1/2}N_1\left(\frac{2\Lambda_3'+1}{2\Lambda_1'+1}\right)^{1/2}U(\Lambda_2\Lambda_3'\Lambda_1\tfrac{1}{2};\Lambda_1'\Lambda_3)\\
&\times\langle(\lambda_1\mu_1)\epsilon_1-3,\Lambda_1';(\lambda_2\mu_2)\epsilon_2\Lambda_2\|(\lambda_3\mu_3)\epsilon_3-3,\Lambda_3'\rangle\\
&+\sum_{\Lambda_2'=\Lambda_2\pm1/2}N_2\left(\frac{2\Lambda_3+1}{2\Lambda_2+1}\right)^{1/2}U(\Lambda_1\Lambda_2'\Lambda_3\tfrac{1}{2};\Lambda_3'\Lambda_2)\\
&\times\langle(\lambda_1\mu_1)\epsilon_1\Lambda_1;(\lambda_2\mu_2)\epsilon_2-3,\Lambda_2'\|(\lambda_3\mu_3)\epsilon_3-3,\Lambda_3'\rangle\Bigg),
\end{aligned}$$

$$N_i=\begin{cases}\sqrt{S(q_i+1)}, & \Lambda_i-\Lambda_i'=\tfrac{1}{2},\\ \sqrt{R(q_i+1)}, & \Lambda_i-\Lambda_i'=-\tfrac{1}{2}.\end{cases} \quad (19)$$

The process is easily realized for small values of $n\equiv[(\lambda_1+\lambda_2-\lambda_3)-2(\mu_1+\mu_2-\mu_3)]/3$. The maximum possible multiplicity is $n+1$, i.e., $\rho_{\max}\leq n+1$. For example, for an allowed coupling with $n=0$, $\rho_{\max}$ must be one and $\langle(\lambda_1\mu_1)\mathrm{HW};(\lambda_2\mu_2)\mathrm{HW}\|(\lambda_3\mu_3)\mathrm{HW}\rangle=1$. For $n=1$, $\rho_{\max}$ may be either one or two. If $\rho_{\max}=2$, the coefficients with $\rho=1$ and $\epsilon_3\Lambda_3=\mathrm{HW}$ are determined via Eqs. (17)–(18) from the result for $\langle(\lambda_1\mu_1)\mathrm{HW};(\lambda_2-1,\mu_2-1)\mathrm{HW}\|(\lambda_3\mu_3)\mathrm{HW}\rangle$. The solution for $\rho=2$ can then be determined from Eq. (20) below. If, on the other hand, $\rho_{\max}=1$, either $(\lambda_1\mu_1)\times(\lambda_2-1,\mu_2-1)\to(\lambda_3\mu_3)$

is allowed and Eq. (20) cannot be used to generate an additional independent solution or $(\lambda_1\mu_1) \times (\lambda_2 - 1, \mu_2 - 1) \to (\lambda_3\mu_3)$ is not allowed and Eq. (20) provides the only solution. For $n = 2$, $\rho_{\max}$ may be either one, two, or three. And in this case it is still possible to generate useful algebraic results. For $n > 2$, however, the recursion process yields unwieldy expressions making the algebraic approach extremely difficult if not impossible. However, from the systematics of the results it is possible to predict a general algebraic expression for $\langle(\lambda_1\mu_1)\mathrm{HW}; (\bar\lambda_2\bar\mu_2)\epsilon_2\Lambda_2 \| (\lambda_3\mu_3)\mathrm{HW}\rangle_\rho$ ($\bar\lambda_2 = \lambda_2 - \eta_{\max} + \rho$, $\bar\mu_2 = \mu_2 - \eta_{\max} + \rho$ implying that ρ is the maximum multiplicity for this coupling) which leads to coefficients that are automatically orthogonal to those obtained via Eq. (13) for the same $\bar\lambda_2, \bar\mu_2$ but lesser ρ. Explicitly, if $\tilde p = \mu - q$ and $\tilde q = \lambda - p$,

$$\langle(\lambda_1\mu_1)\mathrm{HW}; (\bar\lambda_2\bar\mu_2)\epsilon_2\Lambda_2 \| (\lambda_3\mu_3)\mathrm{HW}\rangle_\rho = NF(\tilde{\tilde p}_2)[G(\tilde{\tilde q}_2)/H(\tilde{\tilde q}_2)]^{1/2},$$

$$F(\tilde{\tilde p}_2) = \begin{cases} 1, & \tilde{\tilde p}_2 = 0, \\ (-1)^{\tilde{\tilde p}_2} \sum_{i=0}^{\tilde{\tilde p}_2} \binom{\tilde{\tilde p}_2}{i} \prod_{j=0}^{\tilde{\tilde p}_2 - 1} f(j), & \tilde{\tilde p}_2 \geq 1, \end{cases}$$

$$f(j) = \begin{cases} (\bar p_2 + j + 1)(\mu_1 + \bar\lambda_2 + \bar\mu_2 - n + j + 2), & j < i, \\ (a + j + 1)(b - j - 1), & j \geq i, \end{cases}$$

$$G(\tilde{\tilde q}_2) = \begin{cases} (\bar\lambda_2 - 2\tilde{\tilde q}_2 + n + 1) \binom{n}{\tilde{\tilde q}_2} \prod_{j=j_1}^{j_2 - 1} g(j), & j_2 > j_1, \\ 1, & j_2 = j_1, \end{cases}$$

$$g(j) = \begin{cases} (a + n - j)(b - n + j)(c + n - j)(d + n - j)(\bar\lambda_2 + \bar\mu_2 - j + 1), & j < \tilde{\tilde q}_2, \\ \bar\mu_2 - n + j + 1, & j \geq \tilde{\tilde q}_2, \end{cases} \tag{20}$$

$$\left.\begin{aligned} j_1 &= \text{minimum } \tilde{\tilde q}_2 \\ j_2 &= \text{maximum } \tilde{\tilde q}_2 \end{aligned}\right\} \text{ for which the coupling } \Lambda_1 + \Lambda_2 = \Lambda_3 \text{ is allowed,}$$

$$H(\tilde{\tilde q}_2) = \binom{n + 1 + \bar\lambda_2 - \tilde{\tilde q}_2}{\bar\lambda_2 - \tilde{\tilde q}_2},$$

$$a = A(\lambda_1/2, \lambda_2/2, \lambda_3/2) - n/2, \quad b = B(\lambda_1/2, \lambda_2/2, \lambda_3/2) + n/2 + 1,$$
$$c = C(\lambda_1/2, \lambda_2/2, \lambda_3/2) - n/2, \quad d = D(\lambda_1/2, \lambda_2/2, \lambda_3/2) - n/2,$$
$$n = [(\lambda_1 + \bar\lambda_2 - \lambda_3) + 2(\mu_1 + \bar\mu_2 - \mu_3)]/3,$$

where N is again the normalization factor. The formula [which is essentially the inverse of Eq. (18)]

$$\langle(\lambda_1\mu_1)\epsilon_1 + 3, \Lambda_1'; (\lambda_2\mu_2)\epsilon_2\,\Lambda_2 \| (\lambda_3\mu_3)\mathrm{HW}\rangle$$
$$= \left(\frac{(2\Lambda_1' + 1)}{(2\Lambda_1 + 1)(2\Lambda_2 + 1)N(\Lambda_1')}\right)^{1/2} \sum_{\Lambda_2' = \Lambda_2 \pm 1/2} \frac{Y(\Lambda_1', \Lambda_2')}{\sqrt{2\Lambda_2' + 1}}$$
$$\times \langle(\lambda_1\mu_1)\epsilon_1\Lambda_1; (\epsilon_2\Lambda_2)\epsilon_2 + 3, \Lambda_2' \| (\lambda_3\mu_3)\mathrm{HW}\rangle,$$

$$Y(\Lambda_1 + \tfrac12, \Lambda_2 + \tfrac12) = -\{S(q_2)[A(\Lambda_1, \Lambda_2, \lambda_3/2) + \tfrac12] \times [B(\Lambda_1, \Lambda_2, \lambda_3/2) + \tfrac12]\}^{1/2},$$

$$Y(\Lambda_1 + \tfrac12, \Lambda_2 - \tfrac12) = +\{R(p_2)[C(\Lambda_1, \Lambda_2, \lambda_3/2) + \tfrac12] \times [D(\Lambda_1, \Lambda_2, \lambda_3/2) + \tfrac12]\}^{1/2},$$

$$Y(\Lambda_1 - \tfrac12, \Lambda_2 + \tfrac12) = -\{S(q_2)[C(\Lambda_1, \Lambda_2, \lambda_3/2) + \tfrac12] \times [D(\Lambda_1, \Lambda_2, \lambda_3/2) + \tfrac12]\}^{1/2},$$

$$Y(\Lambda_1 - \tfrac12, \Lambda_2 - \tfrac12) = -\{R(p_2)[A(\Lambda_1, \Lambda_2, \lambda_3/2) + \tfrac12] \times [B(\Lambda_1, \Lambda_2, \lambda_3/2) + \tfrac12]\}^{1/2},$$

$$N(\Lambda_1 + \tfrac12) = S(q_1),$$
$$N(\Lambda_1 - \tfrac12) = R(p_1), \tag{21}$$

can be used to generate coefficients with $\epsilon_1\Lambda_1 \neq \mathrm{HW}$ recursively. Note that Eq. (20) is valid for all ρ; it can be used to provide the starting coefficients for the recursion process.

The computational algorithm is then clear: Neglecting normalization factors, for each $\rho = 1, 2, \ldots, \rho_{\max}$, (i) start with the $\langle(\lambda_1\mu_1)\mathrm{HW}; (\bar\lambda_2\bar\mu_2)\epsilon_2\Lambda_2 \| (\lambda_3\mu_3)\mathrm{HW}\rangle_\rho$ of Eq. (20) and use Eq. (21) to generate the $\langle(\lambda_1\mu_1)\epsilon_1\Lambda_1; (\bar\lambda_2\bar\mu_2)\mathrm{HW} \| (\lambda_3\mu_3)\mathrm{HW}\rangle_\rho$, (ii) make use of Eq. (17) to generate the $\langle(\lambda_1\mu_1)\epsilon_1\Lambda_1; (\lambda_2\mu_2)\mathrm{HW} \| (\lambda_3\mu_3)\mathrm{HW}\rangle_{\tilde\rho}$ from the $\langle(\lambda_1\mu_1)\epsilon_1\Lambda_1; (\bar\lambda_2\bar\mu_2)\mathrm{HW} \| (\lambda_3\mu_3)\mathrm{HW}\rangle_\rho$, and (iii) obtain the $\langle(\lambda_1\mu_1)\epsilon_1\Lambda_1; (\lambda_2\mu_2)\epsilon_2\Lambda_2 \| (\lambda_3\mu_3)\mathrm{HW}\rangle_{\tilde\rho}$ by using Eq. (18) to step the $\epsilon_2\Lambda_2$ labels. Then (iv) use Eq. (8) with $\alpha_3 = \epsilon_3\Lambda_3 = \mathrm{HW}$ to orthonormalize the resultant coefficients in the increasing order $\rho = 1, 2, \ldots, \rho_{\max}$ and, depending upon need, and (v) obtain the $\langle(\lambda_1\mu_1)\epsilon_1\Lambda_1; (\lambda_2\mu_2)\epsilon_2\Lambda_2 \| (\lambda_3\mu_3)\epsilon_3\Lambda_3\rangle_\rho$ by using Eq. (19) to step the $\epsilon_3\Lambda_3$ labels.

The process serves to define $SU_3 \supset SU_2 \times U_1$ Wigner coefficients to within an overall phase. The simplest and most natural way for fixing the phase is to take all the normalization factors involved in the process to be positive, and we adopt this convention. This is very different from the ordinary procedure in which a particular coefficient is assigned to be positive for each mode of coupling, i.e., each ρ-label.[21] With the current approach, however, it is difficult to predict the sign of each individual coefficient, making *a priori* introduction of the ordinary convention practically impossible. Of course, the technique outlined above allows the ordinary convention to be introduced *a posteriori* during the orthonormalization process. And such a choice reflects it-

self in the ρ-dependence of the symmetry properties of the Wigner coefficients (see Sec. 4).

It is interesting to note the effect of changing the order of the coupling in Eq. (11),

$$K^{(\lambda_2\mu_2)}_{\epsilon_2\Lambda_2 M_{\Lambda_2}}(\tilde{\rho}) = [K^{(11)} \times K^{(\lambda_2-1,\mu_2-1)}(\rho)]^{(\lambda_2\mu_2)}_{\epsilon_2\Lambda_2 M_{\Lambda_2}}$$
$$= \sum_{\epsilon\Lambda\Lambda_2' M_\Lambda} \langle(\lambda_2-1,\mu_2-1)\epsilon_2'\Lambda_2' M_{\Lambda_2}'; (11)\epsilon\Lambda M_\Lambda \,|\, (\lambda_2\mu_2)\epsilon_2\Lambda_2 M_{\Lambda_2}\rangle$$
$$\times K^{(11)}_{\epsilon\Lambda M_\Lambda} K^{(\lambda_2-1,\mu_2-1)}_{\epsilon_2'\Lambda_2' M_{\Lambda_2}'}(\rho). \tag{11'}$$

For this form, the result corresponding to Eq. (13) is

$$\langle(\lambda_1\mu_1)\epsilon_1\Lambda_1; (\lambda_2\mu_2)\epsilon_2\Lambda_2 \,\|\, (\lambda_3\mu_3)\epsilon_3\Lambda_3\rangle_{\tilde{\rho}}$$
$$= \langle(\lambda_3\mu_3)\,\|\,K^{(\lambda_2\mu_2)}(\tilde{\rho})\,\|\,(\lambda_1\mu_1)\rangle^{-1}\langle(\lambda_3\mu_3)\,\|\,K^{(11)}\,\|\,(\lambda_3\mu_3)\rangle$$
$$\times \langle(\lambda_3\mu_3)\,\|\,K^{(\lambda_2-1,\mu_2-1)}(\rho)\,\|\,(\lambda_1\mu_1)\rangle$$
$$\times \sum_{\epsilon\Lambda\Lambda_2'\Lambda_3'} \langle(\lambda_2-1,\mu_2-1)\epsilon_2'\Lambda_2'; (11)\epsilon\Lambda \,\|\, (\lambda_2\mu_2)\epsilon_2\Lambda_2\rangle$$
$$\times \langle(\lambda_3\mu_3)\epsilon_3'\Lambda_3'; (11)\epsilon\Lambda \,\|\, (\lambda_3\mu_3)\epsilon_3\Lambda_3\rangle_{\rho=1}$$
$$\times \langle(\lambda_1\mu_1)\epsilon_1\Lambda_1; (\lambda_2-1,\mu_2-1)\epsilon_2'\Lambda_2' \,\|\, (\lambda_3\mu_3)\epsilon_3'\Lambda_3'\rangle_\rho$$
$$\times U(\Lambda_1\Lambda_2'\Lambda_3\Lambda; \Lambda_3'\Lambda_2). \tag{13'}$$

The choice $\epsilon_1\Lambda_1 = \text{HW}$ and $\epsilon_2\Lambda_2 = \text{LW}$ rather than $\epsilon_2\Lambda_2 = \text{HW}$ and $\epsilon_3\Lambda_3 = \text{HW}$ can then be used to obtain a recursion relationship analogous to Eq. (17).

B. SU_3 Racah coefficients

A straightforward generalization of the relationships between SU_2 unitary recoupling coefficients and SU_2 Wigner coefficients leads to the corresponding relationships between SU_3 unitary recoupling (Racah or U functions) and SU_3 Wigner coefficients.[19] The most practical of these relationships for evaluating recoupling coefficients in terms of known Wigner coefficients is

$$\sum_{\rho_{1,23}} \langle(\lambda_1\mu_1)\epsilon_1\Lambda_1; (\lambda_{23}\mu_{23})\epsilon_{23}\Lambda_{23}\|(\lambda\mu)\epsilon\Lambda\rangle_{\rho_{1,23}}$$
$$\times U(\lambda_1\mu_1)(\lambda_2\mu_2)(\lambda\mu)(\lambda_3\mu_3); (\lambda_{12}\mu_{12})\rho_{12},$$
$$\times \rho_{12,3}(\lambda_{23}\mu_{23})\rho_{23}, \rho_{1,23})$$
$$= \sum_{\epsilon_2\Lambda_2\Lambda_3\Lambda_{12}} \langle(\lambda_1\mu_1)\epsilon_1\Lambda_1; (\lambda_2\mu_2)\epsilon_2\Lambda_2 \,\|\, (\lambda_{12}\mu_{12})\epsilon_{12}\Lambda_{12}\rangle_{\rho_{12}}$$
$$\times \langle(\lambda_{12}\mu_{12})\epsilon_{12}\Lambda_{12}; (\lambda_3\mu_3)\epsilon_3\Lambda_3 \,\|\, (\lambda\mu)\epsilon\Lambda\rangle_{\rho_{12,3}}$$
$$\times \langle(\lambda_2\mu_2)\epsilon_2\Lambda_2; (\lambda_3\mu_3)\epsilon_3\Lambda_3 \,\|\, (\lambda_{23}\mu_{23})\epsilon_{23}\Lambda_{23}\rangle_{\rho_{23}}$$
$$\times U(\Lambda_1\Lambda_2\Lambda\Lambda_3; \Lambda_{12}\Lambda_{23}). \tag{22}$$

Fixing $\epsilon_1\Lambda_1 = \text{HW}$ and $\epsilon\Lambda = \text{HW}$ in this expression while letting Λ_{23} run over its range of allowed values yields a set of simultaneous equations the solution of which is the required U functions. Note that the choice $\epsilon_1\Lambda_1 = \text{HW}$ and $\epsilon\Lambda = \text{HW}$ makes it possible to evaluate all but one of the Wigner coefficients in Eq. (22) through Eqs. (17)–(18); the other requires Eq. (19) in addition.

The sum on the right-hand side of Eq. (13) can with the help of Eq. (22) be identified (apart from orthogonality) as simply

$$\sum_{\rho_D} \langle(\lambda_1\mu_1)\epsilon_1\Lambda_1; (\lambda_2\mu_2)\epsilon_2\Lambda_2\|(\lambda_3\mu_3)\epsilon_3\Lambda_3\rangle_{\rho_0} U((\lambda_1\mu_1)(11)(\lambda_3\mu_3)$$
$$\times (\lambda_2-1,\mu_2-1); (\lambda_1\mu_1)\rho_A = 1, \rho_B(\lambda_2\mu_2)\rho_C = 1, \rho_D).$$

This is a direct consequence of the special character of the couplings involved in the product tensors of Eq. (11). More general couplings would, by analogy with SU_2, require a 9-$(\lambda\mu)$ symbol.[22] The recursion formula (13) could therefore, in retrospect, be obtained from Eq. (22) by requiring $U((\lambda_1\mu_1)(11)(\lambda_3\mu_3)(\lambda_2-1,\mu_2-1); (\lambda_1\mu_1)\rho_A = 1, \rho_B(\lambda_2\mu_2)\rho_C = 1, \rho_D) = 0$ for $\rho_D \neq \rho_B$. And indeed, this suggests a simple method by which the techniques developed in this article may be generalized to other group structures. Note that the orthonormalization process, if carried out in the increasing order $\rho = 1, 2, \ldots, \rho_{\max}$, maintains the zero value of the U function for $\rho_D > \rho_B$. Consequently,

$$U((\lambda_1\mu_1)(11)(\lambda_3\mu_3)(\lambda_2-1,\mu_2-1); (\lambda_1\mu_1)\rho_A = 1,$$
$$\rho_B(\lambda_2\mu_2)\rho_C = 1, \rho_D) = 0 \quad \text{for } \rho_D > \rho_B. \tag{23}$$

This result also follows from property (15) and is a direct consequence of the Biedenharn and Louck prescription for specifying the outer multiplicity.

C. $SU_3 \supset R_3$ Wigner coefficients

The coefficients which effect the transformation between the $\epsilon\Lambda M_\Lambda$ and $\mathcal{K}LM$ schemes are known.[23] [The choice made in Eq. (3) requires that an additional factor of $2L+1$ be included in evaluating Eq. (35) of Ref. 23. In addition, including the phase factor $(i)^{n_1+n_3}$ in the definition of $|G\rangle$ makes the coefficient real.] Explicitly, if

$$|(G_E)\mathcal{K}LM\rangle = \sum_g \langle G|(G_E)\mathcal{K}LM\rangle\,|G\rangle, \tag{24}$$

then

$$\langle G|(G_E)\mathcal{K}_i LM\rangle = \sum_{j\leq i} O_{ij}\langle G|(G_E)K_j LM\rangle, \tag{25}$$

where O_{ij} is the orthonormalization matrix of Eq. (6) and $\langle G|(G_E)KLM\rangle$ is the inner product of a state $|G\rangle$ [defined by Eq. (2)] with a state $|(G_E)KLM\rangle$ [defined by Eqs. (3)–(4)]. The parameter g in Eq. (24) is used to denote the subgroup labels $(g_{12}, g_{22}, g_{11} \sim p, q, r)$ of G. In terms of summation ($K \to M, M \to M'$ for reasons of symmetry),

$$\langle G|(G_{\text{HW}})MLM'\rangle = C\sum_{\gamma=0}^{p}\binom{p}{\gamma} S_1(M_\Lambda'\Lambda' N_\Lambda' M')\, S_1(N_\Lambda\Lambda M_\Lambda = \Lambda M) S_2(\kappa' k\kappa = kM'LM),$$

$$C = (-1)^{L-p}\left(\frac{2L+1}{2^p}\right)^2\left[\binom{\lambda}{p}\binom{\mu}{q}\binom{\lambda+\mu+1}{q}\binom{2L}{L-M}\Big/\binom{2L}{L-M'}\binom{2\Lambda'}{\Lambda'+M_\Lambda'}\binom{p+\mu+1}{q}\right]^{1/2},$$

$$S_1(M_\Lambda\Lambda N_\Lambda M) = \sum_\alpha \binom{\Lambda-N_\Lambda}{\alpha}\binom{\Lambda+N_\Lambda}{\Lambda-M_\Lambda-\alpha}\sum_\beta(-1)^\beta\binom{2\Lambda-M_\Lambda-N_\Lambda-2\alpha}{\beta}\binom{2\alpha+M_\Lambda+N_\Lambda}{\Lambda+M/2-\beta},$$

$$S_2(\kappa' k\kappa M'LM)$$
$$= \frac{1}{2k+L+1}\sum_\alpha(-1)^\alpha\binom{L-M}{\alpha}\binom{L+M}{L-M'-\alpha}\sum_\beta(-1)^\beta\binom{k+\kappa'}{\beta}\Big/\binom{2k+L}{q+\Lambda+M/2+\Lambda'+M'/2+\alpha-\beta-\gamma}$$

$$\begin{aligned}
&\Lambda' = (p+\mu-q)/2, && \Lambda = \lambda/2, && k = (\lambda+\mu-\gamma)/2,\\
&M_\Lambda' = r-\Lambda', && M_\Lambda = \Lambda, && \kappa = k,\\
&N_\Lambda' = p-\gamma-\Lambda', && N_\Lambda = \Lambda-\gamma, && \kappa' = k-(\lambda-p+\mu-q).
\end{aligned} \tag{26}$$

The corresponding expression for $\langle G|(G_{LW})KLM\rangle$ can be obtained by conjugation (see Sec. 4). Note in particular that the overlap of two projected states required for a determination of the orthonormalization matrix O_{ij}, is given by

$$\langle (G_E)K'LM|(G_E)KLM\rangle = \langle G_E|(G_E)KLK'\rangle. \tag{27}$$

Since $|G\rangle$ differs from $|(\lambda\mu)\epsilon\Lambda M_\Lambda\rangle$ by at most an n_i-dependent phase factor and $n_i^{(1)} + n_i^{(2)} = n_i^{(3)}$, it is convenient to write

$$\langle G_1; G_2|G_3\rangle_\rho = \langle (\lambda_1\mu_1)\epsilon_1\Lambda_1 M_{\Lambda_1}; (\lambda_2\mu_2)\epsilon_2\Lambda_2 M_{\Lambda_2}|(\lambda_3\mu_3)\epsilon_3\Lambda_3 M_{\Lambda_3}\rangle_\rho. \tag{28}$$

The $SU_3 \supset R_3$ Wigner coefficients are then given by

$$\langle (G_{1E})\mathcal{K}_1L_1M_1; (G_{2E})\mathcal{K}_2L_2M_2|(G_{3E})\mathcal{K}_3L_3M_3\rangle_\rho = \sum_{g_1g_2g_3} \langle G_1|(G_{1E})\mathcal{K}_1L_1M_1\rangle\langle G_2|(G_{2E})\mathcal{K}_2L_2M_2\rangle \times \langle G_3|(G_{3E})\mathcal{K}_3L_3M_3\rangle\langle G_1; G_2|G_3\rangle_\rho. \tag{29}$$

An expression which is more convenient to evaluate from a computational point of view may be obtained by directly expanding the inner product

$$\langle (G_{1E})\mathcal{K}_1L_1M_1; (G_{2E})\mathcal{K}_2L_2M_2|(G_{3E})K_3L_3M_3\rangle_\rho = \langle (G_{1E})\mathcal{K}_1L_1M_1; (G_{2E})\mathcal{K}_2L_2M_2|P^{L_3}_{M_3K_3}|G_{3E}\rangle_\rho. \tag{30}$$

Making use of the fact that $R_3(\Omega) = R_1(\Omega)R_2(\Omega)$, the effect of the projection operator acting to the left can be determined. Integrating over Euler angles by means of the Clebsch–Gordan series for rotation matrices then leads to the result

$$\langle (G_{1E})\mathcal{K}_1L_1M_1; (G_{2E})\mathcal{K}_2L_2M_2|(G_{3E})K_3L_3M_3\rangle_\rho = \sum_{\substack{g_1g_2\\ M_1'(M_2')}} \langle L_1M_1'; L_2M_2'|L_3K_3\rangle\langle G_1|(G_{1E})\mathcal{K}_1L_1M_1'\rangle \times \langle G_2|(G_{2E})\mathcal{K}_2L_2M_2'\rangle\langle G_1; G_2|G_{3E}\rangle_\rho \times \langle L_1M_1; L_2M_2|L_3M_3\rangle. \tag{31}$$

Applying Eq. (5) to the 3-space yields the required $SU_3 \supset R_3$ Wigner coefficients. Note that the summation in this case is only over $SU_3 \supset SU_2 \times U_1$ Wigner coefficients of the type $G_3 = G_{3E}$, i.e., those which can be evaluated through Eqs. (17)–(18) without the use of Eq. (19). Clearly a factorization into the product of a reduced $SU_3 \supset R_3$ Wigner coefficient and an ordinary Wigner coefficient in R_3 space is possible. Note that it is unnecessary and indeed redundant to fix the phase for the $SU_3 \supset R_3$ Wigner coefficients independently of that already chosen for the $SU_3 \supset SU_2 \times U_1$ reduction. The orthonormality of the transformation coefficients between the two schemes guarantees a unique solution. In effect the choice is made by selecting positive roots in Eq. (6).

4. CONJUGATION AND SYMMETRY PROPERTIES

Since the $SU_3 \supset R_3$ reduction is linked to the $SU_3 \supset SU_2 \times U_1$ reduction via the transformation coefficients of Eq. (24), it suffices to make a determination of the conjugation relationship and all symmetry properties for the $SU_3 \supset SU_2 \times U_1$ reduction only. The corresponding $SU_3 \supset R_3$ results follow from known relationships among the transformation coefficients between the two schemes.

A. State conjugation

The transformation coefficients $\langle G|(G_E)\mathcal{K}LM\rangle$ are the elements of a real unitary (hence orthogonal) matrix if $|G\rangle = (i)^{n_1+n_3}|(h_{ij})\rangle$, where the $|(h_{ij})\rangle$ are states of the type defined by Moshinsky in terms of polynomials in creation operators acting on the vacuum.[24] [The choice $(i)^{n_2}|(h_{ij})\rangle$ as suggested in Ref. 23 is also acceptable. In this case, however, the states would not transform according to conventional phasing under the $\mathcal{RT}$ operation (π-rotation about the 2 axis × time reversal) as defined by Bohr and Mottelson.[25]] The results given in Appendix A2 of Ref. 19 for the adjoint irreducible representation can then be used to show that[26]

$$|G\rangle^* = (-1)^{p-r}|\tilde{G}\rangle,$$
$$\tilde{\lambda} = \mu, \quad \tilde{\mu} = \lambda,$$
$$\left.\begin{aligned} \tilde{p} &= \mu - q \\ \tilde{q} &= \lambda - p \\ \tilde{r} &= p + \mu - q - r \end{aligned}\right\} \Rightarrow \begin{aligned} \tilde{\epsilon} &= -\epsilon \\ \tilde{\Lambda} &= \Lambda \\ \tilde{M}_\Lambda &= -M_\Lambda \end{aligned} \tag{32}$$

Note that $p - r = \frac{1}{3}(\lambda - \mu) - \frac{1}{6}\epsilon - M_\Lambda = \tilde{r} - \tilde{p}$. The sign of M_Λ differs from that of Hecht due to the choice $M_\Lambda = r - \Lambda$ of Eq. (1). [This choice allows the more natural correspondence $(zxy) \sim (312)$ rather than $(zxy) \sim (321)$ to be made between body-fixed axes x, y, z and the i, j labels of the Gel'fand scheme.] For $G = G_E$ Eq. (32) implies that in addition to λ and μ interchanging roles $\bar{I} = 1 - I$ and $\bar{J} = 1 - J$, where I and J are as defined by Table I; that is, under conjugation HW→LW and LW→HW.

To discover the conjugation properties of the $|(G_E)\mathcal{K}LM\rangle$ it suffices to know in addition to Eq. (32) the symmetry properties of the $\langle G|(G_E)KLM\rangle$. By straightforward but tedious substitution it can be shown that for the inner product of $|G'\rangle$ with a state $|(G)MLM'\rangle$ [defined by Eq. (3)],

$$\begin{aligned}
&1.\ \langle G'|(G)MLM'\rangle^* = \langle G'|(G)MLM'\rangle,\\
&2.\ \langle G'|(G) - M, L, -M'\rangle = (-1)^{n_2-n_2'+M-M'}\langle G'|(G)MLM'\rangle,\\
&3.\ \langle G'|(G)MLM'\rangle = (-1)^{n_3-n_3'+M-M'}\langle G|(G')M'LM\rangle,\\
&4.\ \langle \tilde{G}'|(\tilde{G})MLM'\rangle = \langle G'|(G)MLM'\rangle,\\
&5A.\ \langle G'|(G)ML, -M'\rangle = (-1)^{n_1'-n_3+L+M}\langle G'|(G)MLM'\rangle,\\
&5B.\ \langle G'|(G) - MLM'\rangle = (-1)^{n_1-n_3'+L+M'}\langle G'|(G)MLM'\rangle,\\
&6A.\ \langle G'(M_\Lambda')|(G)MLM'\rangle = (-1)^{\Lambda'+M'/2}\langle G'(-M_\Lambda')|(G)MLM'\rangle,\\
&6B.\ \langle G'|(G(M_\Lambda))MLM'\rangle = (-1)^{\Lambda+M/2}\langle G'|(G(-M_\Lambda))MLM'\rangle.
\end{aligned} \tag{33}$$

Since $\mathcal{K} = K + 2n$ where n is integral, the symmetries apply directly to the $\langle G|(G_E)\mathcal{K}LM\rangle$ as well as the $\langle G|(G_E)KLM\rangle$. Property 6 together with property 1 insures that the $\langle G'|(G)MLM'\rangle$ vanish for either $2\Lambda' + M'$ or $2\Lambda + M$ odd. Properties 1, 5A, 4 can then be used to show that

$$|(G_E)\mathcal{K}LM\rangle^* = (-1)^{\lambda+\mu+L-M}|(\tilde{G}_E)\mathcal{K}L, -M\rangle. \tag{34}$$

Note that $G_E = G_{HW}$ (G_{LW}) implies Eq. (4a) [Eq. (4b)] applies on the left whereas Eq. (4b) [Eq. (4a)] applies on the right. But since λ and μ also interchange roles, $\mathcal{K}$ is left invariant.

B. Symmetry properties

In Sec. 3 a prescription is given for a unique determination, including phase, of all $SU_3 \supset SU_2 \times U_1$ Wigner coefficients. In terms of $\varphi = \lambda_1 + \lambda_2 - \lambda_3 + \mu_1 + \mu_2 - \mu_3$ which is even or odd as $(\lambda_1 + \lambda_2 - \lambda_3 - \mu_1 - \mu_2 + \mu_3)/3 = p_1 - r_1 + p_2 - r_2 - p_3 + r_3$ is even or odd, the corresponding symmetry properties are:

Symmetry Properties of the $SU_3 \supset SU_2 \times U_1$ Wigner Coefficients

$$1A.\ \langle G_1; G_2 | G_3\rangle_\rho = (-1)^{\varphi+p_2-r_2}\sqrt{\dim(\lambda_3\mu_3)/\dim(\lambda_1\mu_1)}\langle G_3; \tilde{G}_2 | G_1\rangle_\rho,$$

$$2A.\ \langle G_1; G_2 | G_3\rangle_\rho = (-1)^{\varphi+\eta_{\max}-\rho}\langle \tilde{G}_1; \tilde{G}_2 | \tilde{G}_3\rangle_\rho,$$

$$3A.\ \langle G_1; G_2 | G_3\rangle = (-1)^{\varphi}\langle G_2; G_1 | G_3\rangle \quad (\eta_{\max}=1 \text{ only}),$$

$$1B.\ \langle G_1; G_2 \| G_3\rangle_\rho = (-1)^{\varphi+\frac{1}{3}(\lambda_2-\mu_2)-\frac{1}{6}\epsilon_{2^+}\Lambda_3-\Lambda_1} \times \sqrt{\dim(\lambda_3\mu_3)(2\Lambda_1+1)/\dim(\lambda_1\mu_1)(2\Lambda_3+1)} \times \langle G_3; \tilde{G}_2 \| G_3\rangle_\rho,$$

$$2B.\ \langle G_1; G_2 \| G_3\rangle_\rho = (-1)^{\varphi+\eta_{\max}-\rho+\Lambda_1+\Lambda_2-\Lambda_3}\langle \tilde{G}_1; \tilde{G}_2 \| \tilde{G}_3\rangle_\rho,$$

$$3B.\ \langle G_1; G_2 \| G_3\rangle_\rho = (-1)^{\varphi+\Lambda_1+\Lambda_2-\Lambda_3}\langle G_2; G_1 \| G_3\rangle \quad (\eta_{\max}=1 \text{ only}). \quad (35)$$

Among these, the most important is Symmetry 1. Expression (20) satisfies this relation, from which it follows that it holds for the coefficients $\langle G_1; \bar{G}_2 | G_3\rangle_\rho$ and $\langle G_3; \bar{G}_2 | G_1\rangle_\rho$. A comparison of the expression for $\langle G_1; G_2 | G_3\rangle_\rho$ given by the right-hand side of Eq. (13) with that for $\langle G_3; \bar{G}_2 | G_1\rangle_\rho$ given by the right-hand side of Eq. (13′) then sufficies by induction to establish the relationship for the general case. The validity of Symmetry 2, apart from phase, is a direct consequence of the symmetric nature of the formulation under the operation of conjugation. The appearance of the phase factor in this case, however, is by no means obvious. The factor $(-1)^\varphi$ is a direct consequence of Eq. (32). But, as already suggested,[26] consistency requires an additional phase, $\xi = \pm 1$. It has been determined that $\langle G_1; \bar{G}_2 | G_3\rangle_\rho = (-1)^\varphi\langle \tilde{G}_1; \tilde{G}_2 | \tilde{G}_3\rangle_\rho$, i.e., $\xi = +1$ for this special variety. The general result, $\xi = (-1)^{\eta_{\max}-\rho}$, then follows from recursion relation (13). An arbitrary resolution of the multiplicity would, in general, require a linear transformation among the ρ-labels on the right-hand side of each of Eqs. (35). The significance of the "canonical" decomposition manifests itself in Symmetry 1 and Symmetry 2, where such a transformation does not appear and the multiplicity label ρ is the same on both sides of the equations. This, however, is not the case for Symmetry 3 because of the unsymmetric treatment of G_1 and G_2 and accounts for the restriction $\eta_{\max}=1$, i.e., multiplicity free couplings only.

Practical considerations may favor adopting a different phase convention.[21] But doing so requires a modification in the phases for the symmetries of Eq. (35). For example, under the convention adopted by Hecht, namely requiring $\langle(\lambda_1\mu_1)\text{LW}; (\lambda_2\mu_2)\epsilon_2\Lambda_{2_{\max}} \| (\lambda_3\mu_3)\text{LW}\rangle_\rho > 0$, the results can be summarized as follows:

Symmetry 1 remains unchanged, Symmetry 2 holds with $\eta_{\max}$ replaced by $\rho_{\max}$, Symmetry 3 is valid for $\rho_{\max}=1$ only. That is, in this particularly simple case all that is required is for $\eta_{\max}$ to be replaced by $\rho_{\max}$ throughout.

The symmetry properties of the $SU_3 \supset R_3$ Wigner coefficients can be obtained from those given above by using the results of Eqs. (33) together with Eq. (34).

Symmetry Properties of the $SU_3 \supset R_3$ Wigner Coefficients

$$1A.\ \langle(G_{1E})\mathcal{K}_1L_1M_1; (G_{2E})\mathcal{K}_2L_2M_2 | (G_{3E})\mathcal{K}_3L_3M_3\rangle_\rho = (-1)^{\varphi+\lambda_2+\mu_2+L_2+M_2}\sqrt{\dim(\lambda_3\mu_3)/\dim(\lambda_1\mu_1)} \times \langle(G_{3E})\mathcal{K}_3L_3M_3; (\tilde{G}_{2E})\mathcal{K}_2L_2, -M_2 | (G_{3E})\mathcal{K}_3L_3M_3\rangle_\rho$$

$$2A.\ \langle(G_{1E})\mathcal{K}_1L_1M_1; (G_{2E})\mathcal{K}_2L_2M_2 | (G_{3E})\mathcal{K}_3L_3M_3\rangle_\rho = (-1)^{\varphi+\eta_{\max}-\rho+L_1+L_2-L_3} \times \langle(\tilde{G}_{1E})\mathcal{K}_1L_1, -M_1; (\tilde{G}_{2E})\mathcal{K}_2L_2, -M_2 | (\tilde{G}_{3E})\mathcal{K}_3L_3, -M_3\rangle_\rho,$$

$$3A.\ \langle(G_{1E})\mathcal{K}_1L_1M_1; (G_{2E})\mathcal{K}_2L_2M_2 | (G_{3E})\mathcal{K}_3L_3M_3\rangle = (-1)^{\varphi} \times \langle(G_{2E})\mathcal{K}_2L_2M_2; (G_{1E})\mathcal{K}_1L_1M_1 | (G_{3E})\mathcal{K}_3L_3M_3\rangle \quad (\eta_{\max}=1 \text{ only}),$$

$$1B.\ \langle(G_{1E})\mathcal{K}_1L_1; (G_{2E})\mathcal{K}_2L_2 \| (G_{3E})\mathcal{K}_3L_3\rangle_\rho = (-1)^{\varphi+\lambda_2+\mu_2+L_1+L_2-L_3} \times \sqrt{\dim(\lambda_3\mu_3)(2L_1+1)/\dim(\lambda_1\mu_1)(2L_3+1)} \langle(G_{3E})\mathcal{K}_3L_3; (\tilde{G}_{2E})\mathcal{K}_2L_2 \| (G_{1E})\mathcal{K}_1L_1\rangle_\rho,$$

$$2B.\ \langle(G_{1E})\mathcal{K}_1L_1; (G_{2E})\mathcal{K}_2L_2 \| (G_{3E})\mathcal{K}_3L_3\rangle_\rho = (-1)^{\varphi+\eta_{\max}-\rho} \times \langle(\tilde{G}_{1E})\mathcal{K}_1L_1; (\tilde{G}_{2E})\mathcal{K}_2L_2 \| (\tilde{G}_{3E})\mathcal{K}_3L_3\rangle_\rho,$$

$$3B.\ \langle(G_{1E})\mathcal{K}_1L_1; (G_{2E})\mathcal{K}_2L_2 \| (G_{3E})\mathcal{K}_3L_3\rangle = (-1)^{\varphi+L_1+L_2-L_3} \times \langle(G_{2E})\mathcal{K}_2L_2; (G_{1E})\mathcal{K}_1L_1 \| (G_{3E})\mathcal{K}_3L_3\rangle \quad (\eta_{\max}=1 \text{ only}). \quad (36)$$

Again, under the convention of Hecht, these relations hold if $\eta_{\max}$ is replaced by $\rho_{\max}$ throughout.

5. CONCLUDING REMARKS

The techniques described above developed as an outgrowth of the need for an advanced SU_3 technology in shell model calculations for light nuclei assuming general two-body effective interactions.[27] Machine codes based on the results are therefore available.[28] They allow a numerical determination of $SU_3 \supset SU_2 \times U_1$ and $SU_3 \supset R_3$ Wigner coefficients as well as SU_3 Racah coefficients to be made for arbitrary couplings and multiplicity.

Although the emphasis in the present article has been on the practical aspects of calculating SU_3 Wigner and Racah coefficients, it is quite possible, and indeed likely, that the build-up process using the group generators can be applied to the Γ_s Wigner operators of Biedenharn and Louck and co-workers for the couplings $(\lambda_1\mu_1) \times (\bar{\lambda}_2\bar{\mu}_2) \to (\lambda_3\mu_3)$, $\rho = 1, 2, \ldots, \rho_{\max}$ to obtain the full set of Wigner operators for the coupling $(\lambda_1\mu_1) \times (\lambda_2\mu_2) \to (\lambda_3\mu_3)$. Because of nonorthogonality, however, it is not clear that a simple interpretation of the structure of the operators in terms of geometrical properties of the so-called arrow patterns will be possible. Nevertheless, since our purpose in the present article is to avoid the luxury of mathematical sophistication the validity of such conjectures must be relegated to a later work.

ACKNOWLEDGMENTS

Thanks are due to nearly all the individuals with whom the authors had contact during the period in which the results reported in this article developed from mere speculation to concrete formulation. Special thanks, however, are due to K. T. Hecht for the expertise he provided in this technical phase of our research into the structure of light nuclei. It is also a pleasure to acknowledge the support of W. C. Parkinson and the cyclotron group at Michigan.

*Work supported by the U.S. National Science Foundation.

† Present address: Department of Physics, University of Rochester, Rochester, New York.

‡ Present address: Department of Physics, Nihon University, Setagaya, Tokyo, Japan.

[1] E. P. Wigner, *Group Theory and Its Application to the Quantum Mechanics of Atomic Spectra*, translated from the German by J. J. Griffin (Academic, New York, 1959).

[2]G. Racah, "Theory of Complex Spectra," in *Quantum Mechanics of Angular Momentum*, edited by L. C. Biedenharn and H. Van Dam (Academic, New York, 1965).
[3]J. P. Elliott, Proc. R. Soc. A **245**, 128, 562 (1958).
[4]Y. Ne'eman, Nucl. Phys. **26,** 222 (1961). M. Gell-Mann, Phys. Rev. **125,** 1067 (1962).
[5]See, for example, references cited by C. K. Chew and R. T. Sharp, Nucl. Phys. B **2,** 697 (1967).
[6]L. C. Biedenharn and J. D. Louck, J. Math. Phys. **13,** 1985 (1972).
[7]L. C. Biedenhard, J. D. Louck, E. Chacón, and M. Ciftan, J. Math. Phys. **13,** 1957 (1972).
[8]E. Chacón, M. Ciftan, and L. C. Biedenharn, J. Math. Phys. **13,** 577 (1972).
[9]J. D. Louck and L. C. Biedenharn, J. Math. Phys. **12,** 173 (1971).
[10]J. D. Louck and L. C. Biedenharn, J. Math. Phys. **11,** 2368 (1970).
[11]J. A. Castilho Alcarás, L. C. Biedenharn, K. T. Hecht, and G. Neeley, Ann. Phys. (N.Y.) **60,** 85 (1970).
[12]J. D. Louck, Am. J. Phys. **38,** 3 (1970) presents a bibliography and synthesis of published results on the theory of tensor operators in the unitary groups.
[13]J. P. Elliott and M. Harvey, Proc. R. Soc. A **272,** 557 (1962).
[14]R. E. Behrends, J. Dreitlin, C. Fronsdal, and B. W. Lee, Rev. Mod. Phys. **34,** 1 (1962). J. J. deSwart, Rev. Mod. Phys. **35,** 916 (1963).
[15]I. M. Gel'fand and M. L. Zeitlin, Dokl. Akad. Nauk SSSR **71,** 825 (1950).
[16]G. Racah, "Group Theory and Spectroscopy" (Lecture Notes, Princeton, 1951). V. Bargmann and M. Moshinsky, Nucl. Phys. **18,** 697 (1960); Nucl. Phys. **23,** 177 (1961). R. T. Sharp and Hans C. Von Baeyer, Nucl. Phys. A. **127,** 513 (1969).
[17]M. Harvey, in *Advances in Nuclear Physics,* edited by M. Baranger and and E. Vogt (Plenum, New York, 1971), Vol. 1.
[18]J. D. Vergados, Nucl. Phys. A **111,** 681 (1968).
[19]K. T. Hecht, Nucl. Phys. **62,** 1 (1965).
[20]A. R. Edmonds, *Angular Momentum in Quantum Mechanics* (Princeton U. P., Princeton, N. J., 1957).
[21]In addition to Refs. 14 and 19 see, for example, M. Moshinsky, Rev. Mod. Phys. **34,** 813 (1962). A. R. Edmonds, Proc. R. Soc. A **268,** 567 (1963).
[22]A. deShalit and I. Talmi, *Nuclear Shell Theory* (Academic, New York, 1963).
[23]J. P. Draayer and S. A. Williams, Nucl. Phys. A 129, 647 (1969).
[24]E. Chacón and M. Moshinsky, Phys. Lett. **23,** 567 (1966); in *Spectroscopic and Group Theoretical Concepts and Methods in Elementary Particle Physics*, edited by F. Bloch, S. G. Cohen, A. deShalit, S. Sambursky, and I. Talmi (North-Holland, Amsterdam, 1968).
[25]A. Bohr and Ben R. Mottelson, *Nuclear Structure* (Benjamin, New York, 1969).
[26]For a coupled system $(\lambda_1\mu_1) \times (\lambda_2\mu_2) \to (\lambda_3\mu_3)$, Eq. (32) cannot simultaneously be applied to all three representations $(\lambda_1\mu_1)$, $(\lambda_2\mu_2)$, and $(\lambda_3\mu_3)$ in a consistent fashion. An additional dependence upon the multiplicity label is required. See, for example, J. J. deSwart, Rev. Mod. Phys. **35,** 916 (1963), Sec. 14 and Sec. 4B of the current article.
[27]J. P. Draayer, Bull. Am. Phys. Soc. **18** (1), 137 (1973). R. D. Ratna Raju, J. P. Draayer, and K. T. Hecht, Nucl. Phys. A 202, 433 (1973).
[28]J. P. Draayer and Yoshimi Akiyama, Comput. Phys. Commun. 5, 405 (1973).

Linear integral transformations generated by the three-dimensional neutron transport kernel

V. C. Boffi and G. Spiga

Laboratorio di Ingegneria Nucleare dell'Università di Bologna, Bologna, Italy
(Received 12 June 1973)

Two theorems are established by which the theory of linear integral transformations in a Lebesgue space $L_p(D)$, $p \geq 1$, can be appropriately extended to solve linear integral equations with kernel of nonfinite double norm with respect to the considered $L_p(D)$. An application of these theorems to a physical problem of three-dimensional neutron transport theory is illustrated.

1. INTRODUCTION

Let $\bar{x}, \bar{y} \equiv (x_1, \ldots, x_n; y_1, \ldots, y_n) \in D$, where D is a Lebesgue measurable domain (of infinite or finite measure) of the Euclidean n-dimensional space R_n, with Lebesgue measure $d\bar{x} = dx_1 \cdots dx_n$.

We consider the linear integral equation

$$f(\bar{x}) = g(\bar{x}) + \int_D k(\bar{x}, \bar{y}) f(\bar{y}) d\bar{y}, \tag{1}$$

where $f(\bar{x})$, $g(\bar{x})$, and $k(\bar{x}, \bar{y})$ are Lebesgue measurable functions on D, D, and $D \times D$, respectively.

It is well known (see, for instance Ref. 1) that the existence and the uniqueness of the Neumann series solution $f(\bar{x})$, of Eq. (1), can be proved by resorting to the theory of linear integral transformations in a Lebesgue space $L_p(D)$ $(p \geq 1)$, provided the free term $g(\bar{x}) \in L_p(D)$ and $k(\bar{x}, \bar{y})$ is an element of the Banach space $N_p(D)$, that is, $k(\bar{x}, \bar{y})$ is of finite double norm $|||K|||_p$ with respect to the same Lebesgue space $L_p(D)$ which $g(\bar{x})$ belongs to. The double norm is defined as[1]

$$|||k|||_p = \| \, \| k(\bar{x}, \bar{y}) \|_q \|_p = \left\{ \int_D \left[\int_D | k(\bar{x}, \bar{y}) |^q d\bar{y} \right]^{p/q} d\bar{x} \right\}^{1/p} \tag{2}$$

for any pair of real numbers $p, q \geq 1$ such that $1/p + 1/q = 1$.

In this paper we study Eq. (1), where the free term $g(\bar{x})$ is still taken to be of class $L_p(D)$, in the anomalous case when the kernel $k(\bar{x}, \bar{y})$ is not any longer of finite double norm with respect to $L_p(D)$. Our theory does not exclude however the case when the kernel $k(\bar{x}, \bar{y})$ is of finite double norm with respect to a Lebesgue space $L_{p'}(D)$, whose index p' is different from the index p of the space $L_p(D)$, which the free term $g(\bar{x})$ belongs to.

The main feature of the theory is to replace the hypothesis that the kernel $k(\bar{x}, \bar{y})$ of Eq. (1) is of finite double norm by imposing on it appropriate sufficient conditions. These conditions constitute the body of two theorems, by which the nature of the transformation

$$Tf = \int_D k(\bar{x}, \bar{y}) f(\bar{y}) d\bar{y}, \tag{3}$$

generated by the kernel $k(\bar{x}, \bar{y})$ of Eq. (1), can be explicitly investigated. Thence it can be shown that the existence, the uniqueness, and other properties of the solution of Eq. (1) can be again established via the Neumann series representation, as usually done in the case when both the free term and the kernel of Eq. (1) belong to Banach spaces of same index p.

We shall present at the end of the paper an application of the theory to a problem drawn from three-dimensional transport theory for monoenergetic neutrons.

2. A FIRST THEOREM

We suppose that the kernel $k(\bar{x}, \bar{y})$ and its adjoint $k^*(\bar{x}, \bar{y}) = k(y, x)$ are of finite double norm with respect to $L_\infty(D)$, that is we have

$$\int_D | k(\bar{x}, \bar{y}) | \, d\bar{x} \leq A < \infty \tag{4a}$$

for almost every $\bar{y} \in D$, and

$$\int_D | k(\bar{x}, \bar{y}) | \, d\bar{y} \leq B < \infty \tag{4b}$$

for almost every $\bar{x} \in D$.

Then the following theorem holds.

Theorem 1: The transformation

$$Tf = \int_D k(\bar{x}, \bar{y}) f(\bar{y}) d\bar{y}, \tag{3}$$

whose kernel satisfies the conditions, Eqs. (4a) and (4b), is a linear integral transformation of $L_p(D)$ into $L_p(D)$, and it is bounded—and therefore continuous—with

$$\| T \|_p \leq A^{1/p} B^{1-1/p}, \tag{5}$$

for any $1 \leq p \leq \infty$.

Proof: We consider first the case that $p = 1$. If $f(\bar{x}) \in L_1(D)$ we get

$$\begin{aligned} \| Tf \|_1 &\leq \int_D d\bar{x} \int_D | k(\bar{x}, \bar{y}) | \, | f(\bar{y}) | \, d\bar{y} \\ &= \int_D | f(\bar{y}) | \, d\bar{y} \int_D | k(\bar{x}, \bar{y}) | \, d\bar{x} \leq A \| f \|_1, \end{aligned} \tag{6}$$

where the change of the order of integration is permissible by Fubini's theorem for measurable nonnegative functions.

We consider then the case that $p = \infty$. If $f(\bar{x}) \in L_\infty(D)$ we get

$$\begin{aligned} \| Tf \|_\infty &= \operatorname*{ess.\,sup}_{x \in D} | Tf | \leq \operatorname*{ess.\,sup}_{x \in D} \int_D | k(\bar{x}, \bar{y}) | \, | f(\bar{y}) | \, d\bar{y} \\ &\leq \| f \|_\infty \operatorname*{ess.\,sup}_{x \in D} \int_D | k(\bar{x}, \bar{y}) | \, d\bar{y} \leq B \| f \|_\infty, \end{aligned} \tag{7}$$

where by the essential supremum we mean the upper bound of a function apart from a set of zero measure.

We turn now to the general case $1 < p < \infty$. If $f(\bar{x}) \in L_p(D)$ and $q > 1$ is a real number such that

$$1/p + 1/q = 1, \tag{8}$$

we rewrite Eq. (3) as

$$| Tf | \leq \int_D | k(\bar{x}, \bar{y}) |^{1/p} | f(\bar{y}) | \, | k(\bar{x}, \bar{y}) |^{1/q} d\bar{y}. \tag{9}$$

We realize that $| k(\bar{x}, \bar{y}) |^{1/p} | f(\bar{y}) |$ is p-summable on D,

as it can be deduced from the preceding case $p = 1$, whereas $|k(\bar{x},\bar{y})|^{1/q}$ is, by definition, q-summable on D.

By Hölder's inequality we thus obtain

$$|Tf| \leq B^{1/q}\left[\int_D |k(\bar{x},\bar{y})|\,|f(\bar{y})|^p d\bar{y}\right]^{1/p} \tag{10a}$$

for almost every $\bar{x} \in D$, and then

$$\begin{aligned}\|Tf\|_p &\leq B^{1/q}\left[\int_D d\bar{x}\int_D |k(\bar{x},\bar{y})|\,|f(\bar{y})|^p d\bar{y}\right]^{1/p}\\ &\leq B^{1/q}\left[\int_D |f(\bar{y})|^p d\bar{y}\int_D |k(\bar{x},\bar{y})|\,d\bar{x}\right]^{1/p}\\ &\leq A^{1/p}B^{1/q}\|f\|_p.\end{aligned} \tag{10b}$$

Combining the result of Eq. (10b) with those of Eqs. (6) and (7) shows that the transformation T is a linear integral transformation of $L_p(D)$ into $L_p(D)$ for any $1 \leq p \leq \infty$.

Furthermore, for $f(\bar{x})$ is arbitrary, from Eqs. (6), (7), (8), and (10b) there follows that T is bounded, according to Eq. (5). As a function of $1 \leq p \leq \infty$, the norm $\|T\|_p$ is always bounded by $A_0 = \max\{A, B\}$.

If the kernel $k(\bar{x},\bar{y})$ of Eq. (1) is self-adjoint, then $A = B$. As a consequence of this theorem we have:

Corollary: Every kernel $k(\bar{x},\bar{y})$ satisfying conditions (4a) and (4b) has the property (P) with respect to $L_p(D)$, with $1 \leq p \leq \infty$, that is

$$\int_D |k(\bar{x},\bar{y})|\,|f(\bar{y})|\,d\bar{y} \in L_p(D) \tag{11}$$

for every $f(\bar{x}) \in L_p(D)$.
The proof of the corollary is trivial.
We recall that[1] the property (P) guarantees the existence of the iterated kernels

$$k_n(\bar{x},\bar{y}) = \int_D k(\bar{x},\bar{z})k_{n-1}(\bar{z},\bar{y})d\bar{z} \qquad (n = 2, 3, \cdots) \tag{12a}$$

with $k_1(\bar{x},\bar{y}) = k(\bar{x},\bar{y})$, each of them having the property (P) with respect to $L_p(D)$, and, furthermore,

$$T^n f = \int_D k_n(\bar{x},\bar{y})f(\bar{y})d\bar{y}. \tag{12b}$$

How Theorem 1 can be generalized is next shown.

3. A SECOND THEOREM

We suppose now that for some real number $p' > 1$ the kernel $k(\bar{x},\bar{y})$ of Eq. (1) is such that

$$\left[\int_D |k(\bar{x},\bar{y})|^{p'} d\bar{x}\right]^{1/p'} \leq A_{p'} < \infty \tag{13a}$$

for almost every $\bar{y} \in D$, and

$$\left[\int_D |k(\bar{x},\bar{y})|^{p'} d\bar{y}\right]^{1/p'} \leq B_{p'} < \infty \tag{13b}$$

for almost every $\bar{x} \in D$.

We notice that, if D is of finite measure, Eqs. (13a) and (13b) are conditions sufficient for the kernel $k(\bar{x},\bar{y})$ and its adjoint to be of finite double norm with respect to $L_{q'}(D)$, with $q' = p'/(p'-1)$.

The following theorem is in order.

Theorem 2: For any real number p such that

$$1/p + 1/p' \geq 1, \tag{14}$$

the transformation

$$Tf = \int_D k(\bar{x},\bar{y})f(\bar{y})d\bar{y}, \tag{3}$$

in which the kernel $k(\bar{x},\bar{y})$ obeys conditions, Eqs. (13a) and (13b), and $f(\bar{x}) \in L_p(D)$, is a linear integral transformation of $L_p(D)$ into $L_r(D)$, with

$$1/r = 1/p + 1/p' - 1. \tag{15}$$

There results also that

$$\|Tf\|_r \leq A_{p'}^{p'/r}B_{p'}^{1-p'/r}\|f\|_p. \tag{16}$$

Proof: We consider first the case that $p' = 1$. Then $r = p$ and Theorem 1 applies.

When $1/p + 1/p' = 1$, $r = \infty$ and $q' = p$. Then

$$\begin{aligned}\|Tf\|_\infty &= \operatorname*{ess\,sup}_{\bar{x}\in D} |Tf|\\ &\leq \operatorname*{ess\,sup}_{\bar{x}\in D}\left[\int_D |k(\bar{x},\bar{y})|^{p'} d\bar{y}\right]^{1/p'}\cdot\left[\int_D |f(\bar{y})|^{q'} d\bar{y}\right]^{1/q'}\\ &\leq B_{p'}\|f\|_p.\end{aligned} \tag{17}$$

When $p' = \infty$, we have $p = 1$ according to Eq. (14). Then the result is still given by Eq. (17), with $p' = \infty$ and $p = 1$.

Thus there remains to consider the case that $1 < p' < \infty$, $1/p + 1/p' > 1$, from which $1 < r < \infty$. We rewrite Eq. (3) as

$$\begin{aligned}|Tf| \leq \int_D &|k(\bar{x},\bar{y})|^{p'/r}|f(\bar{y})|^{p/r}\\ &|k(\bar{x},\bar{y})|^{p'(1/p'-1/r)}|f(\bar{y})|^{p(1/p-1/r)}d\bar{y}.\end{aligned} \tag{18}$$

We realize that $|k(\bar{x},\bar{y})|^{p'/r}|f(\bar{y})|^{p/r}$ is r-summable on D by virtue of Theorem 1, whereas $|k(\bar{x},\bar{y})|^{p'(1/p'-1/r)}$ and $|f(\bar{y})|^{p(1/p-1/r)}$ are, by definition, $(1/p' - 1/r)^{-1}$-summable and $(1/p - 1/r)^{-1}$-summable on D, respectively. Now

$$\frac{1}{r} + \frac{1}{p'} - \frac{1}{r} + \frac{1}{p} - \frac{1}{r} = 1 \tag{19}$$

as follows from Eq. (15).

Then, by a two-fold application of Hölder's inequality we get

$$\begin{aligned}|Tf| \leq B_{p'}^{p'(1/p'-1/r)}\|f\|_p^{p(1/p-1/r)}&\\ \times\left[\int_D |k(\bar{x},\bar{y})|^{p'}|f(\bar{y})|^p d\bar{y}\right]^{1/r}&\end{aligned} \tag{20a}$$

for almost every $\bar{x} \in D$, and then

$$\begin{aligned}\|Tf\|_r &\leq B_{p'}^{p'(1/p'-1/r)}\|f\|_p^{p(1/p-1/r)}\\ &\left[\int_D |f(\bar{y})|^p d\bar{y}\int_D |k(\bar{x},\bar{y})|^{p'} d\bar{x}\right]^{1/r}\\ &\leq A_{p'}^{p'/r}B_{p'}^{1-p'/r}\|f\|_p.\end{aligned} \tag{20b}$$

Combining the result of Eq. (20b) with those expressed by Eqs. (17) and (18) proves the theorem.

Examples of known kernels which satisfy conditions, Eqs. (13a) and (13b), are the following ones:

(i) The convolution kernels, that is, kernels of the type

$$k(\bar{x},\bar{y}) = k_0(\bar{x} - \bar{y}), \tag{21a}$$

where $k_0(\bar{x})$ is a function p'-summable on D, being D the entire space R_n.

In fact,

$$\left[\int_{R_n} |k_0(\bar{x}-\bar{y})|^{p'} \frac{d\bar{x}}{d\bar{y}}\right]^{1/p'} = \left[\int_{R_n} |k_0(\bar{z})|^{p'} d\bar{z}\right]^{1/p'} = \|k_0\|_{p'}. \tag{21b}$$

We observe then that the convolution kernels are not of finite double norm.

Theorems for the existence of convolution integrals as

$$K_0 f = \int_{R_n} k_0(\bar{x}-\bar{y}) f(\bar{y}) d\bar{y} \tag{21c}$$

can be found in Ref. 2 for various specializations of the functional spaces which $k_0(\bar{x})$ and $f(\bar{x})$ may belong to.

(ii) The kernels with weak singularity, that is, kernels of the type

$$k(\bar{x},\bar{y}) = t(\bar{x},\bar{y})/|\bar{x}-\bar{y}|^{\alpha}. \tag{22}$$

Now D is a domain of finite diameter Δ, $t(\bar{x},\bar{y})$ is a function which is bounded almost everywhere on $D \times D$, with $|t(\bar{x},\bar{y})| \leq C$, and α is a real number smaller than n. In fact,

$$\left[\int_D |k(\bar{x},\bar{y})|^{p'} \frac{d\bar{x}}{d\bar{y}}\right]^{1/p'} \leq C\left[\int_D \frac{1}{|\bar{x}-\bar{y}|^{\alpha p'}} \frac{d\bar{x}}{d\bar{y}}\right]^{1/p'} \leq C|S|^{1/p'}\left[\int_0^{\Delta} r^{n-1-\alpha p'} dr\right]^{1/p'}. \tag{23a}$$

where $|S| = 2\pi^{n/2}/\Gamma(n/2)$ is the surface area of the hypersphere S with radius unity. If now $n - 1 - \alpha p' > -1$, that is, $p' < n/\alpha$, then Eq. (23a) reduces to

$$\left[\int_D |k(\bar{x},\bar{y})|^{p'} \frac{d\bar{x}}{d\bar{y}}\right]^{1/p'} \leq C|S|^{1/p'}\left[\frac{\Delta^{n-\alpha p'}}{n-\alpha p'}\right]^{1/p'}, \tag{23b}$$

which shows that the kernels of the type as in Eq. (22) belong to the class of kernels satisfying conditions (13a) and (13b) when

$$1 \leq p' < n/\alpha. \tag{23c}$$

4. AN EXAMPLE

As an example, we pass now to consider a three-dimensional problem of monoenergetic neutron transport theory.

In this problem the unknown $f(\bar{x})$ of Eq. (1) is the neutron total flux in a body occupying a measurable domain $D \subset R_3$ of finite diameter Δ.

If the material the body is made of is inhomogeneous, that is, its total and scattering-fission macroscopic cross sections Σ and Σ_{sf}, respectively, are bounded real nonnegative measurable functions of the point $\bar{y} \in D$ at which neutrons are isotropically produced by scattering and fission, then for the kernel $k(\bar{x},\bar{y})$ of Eq. (1) we have

$$k(\bar{x},\bar{y}) = \Sigma_{sf}(\bar{y}) \frac{e^{-\tau(\bar{x},\bar{y})}}{4\pi|\bar{x}-\bar{y}|^2}, \tag{24}$$

where

$$\tau(\bar{x},\bar{y}) = \int_0^{|\bar{x}-\bar{y}|} \Sigma\left(\bar{x} - \frac{\bar{x}-\bar{y}}{|\bar{x}-\bar{y}|}u\right) du \tag{24'}$$

is the optical distance between $\bar{x}$ and $\bar{y}$.

The unsymmetric function $k(\bar{x},\bar{y})$, Eq. (24), represents thus a kernel with weak singularity, according to the definition of Eq. (22).

For the free term $g(\bar{x})$ of Eq. (1) we may write

$$g(\bar{x}) = \int_D Q_1 \delta(\bar{y}-\bar{x}_1) \frac{e^{-\tau(\bar{x},\bar{y})}}{4\pi|\bar{x}-\bar{y}|^2} d\bar{y} = Q_1 \frac{e^{-\tau(\bar{x},\bar{x}_1)}}{4\pi|\bar{x}-\bar{x}_1|^2}, \tag{25}$$

if we refer to the case when an isotropic delta-like source of intensity Q_1 is localized at the point $\bar{x}_1 \in D$. We observe that $g(\bar{x})$ is a real positive measurable function on D, and it is bounded in any subset of D which does not include the point $\bar{x}_1$.

The problem [Eqs. (1), (24) and (25)], which accounts for the mathematical constraints due to both the nuclear heterogeneity of the body and the singularity of the source, has not been investigated up to now in all its implications. How the theory of Secs. 2 and 3 succeeds in treating this typical problem of neutron transport is now shown. It is first recognized that the free term $g(\bar{x})$, Eq. (25), belongs to $L_p(D)$, with $1 \leq p < 3/2$, whereas the kernel $k(\bar{x},\bar{y})$, Eq. (24), belongs to $N_{p'}(D)$ with $p' > 3$, as it is of finite double norm $|||K|||_{p'}$, Eq. (2), with respect to $L_{p'}(D)$ with $p' > 3$. (For the details of these calculations compare Ref. 3).

We are thus led to consider the case—mentioned in the Introduction—when the free term and the kernel of Eq. (1) belong to Banach spaces with different indices p and p'. If p and p' are taken to be such that $1/p + 1/p' = 1$, the spaces $L_p(D)$ and $L_{p'}(D)$ are then complementary.

The situation now to be faced is that the application of the transformation T with kernel $k(\bar{x},\bar{y}) \in N_{p'}(D)$ to a function $h(\bar{x}) \in L_p(D)$ no longer guarantees in general that the resulting function will still be of class $L_p(D)$. The solution of Eq. (1) via Neumann series representation then becomes impracticable.

But this difficulty can be overcome by straightforward application of Theorems 1 and 2. We begin with Theorem 1. The kernel $k(\bar{x},\bar{y})$, Eq. (24), satisfies all the hypotheses of Theorem 1. In fact, by overestimating the function $k(\bar{x},\bar{y})$ by means of the inequalities

$$\Sigma_{sf}(\bar{x}) \leq \Sigma_{sf}^{max} < \infty, \quad \Sigma(\bar{x}) \geq \Sigma_{min} \geq 0, \tag{26}$$

we get

$$A = B = \Sigma_{sf}^{max} \frac{1 - e^{-\Sigma_{min}\Delta}}{\Sigma_{min}} \tag{27a}$$

for $\Sigma_{min} > 0$, and

$$A = B = \Sigma_{sf}^{max}\Delta \tag{27b}$$

for $\Sigma_{min} = 0$. There results thus for the norm of the transformation T, defined by Eq. (3),

$$\|T\|_p \leq A \tag{28}$$

for any $1 \leq p \leq \infty$.

Then Tg is of the same class $L_p(D)$, with $1 \leq p < 3/2$, which $g(\bar{x})$, Eq. (25), belongs to. The same is true for the general term $T^n g$, with $n > 1$, where T^n denotes the linear integral transformation defined by Eqs. (12).

We can then build up the Neumann series associated to Eq. (1), namely

$$f^*(\bar{x}) = g(\bar{x}) + \sum_{n=1}^{\infty} T^n g, \tag{29}$$

whose convergence in norm of $L_p(D)$, for any index $1 \leq p < 3/2$, is guaranteed, according to the Riesz-Fisher theorem, by the completeness of the Lebesgue spaces, once the Cauchy condition

$$\lim_{k,l\to\infty} \| \sum_{n=l+1}^{k} T^n g \|_p = 0 \tag{30}$$

is satisfied.

By using well-known inequalities, Eq. (30) can be rewritten as

$$\begin{aligned} \lim_{k,l\to\infty} \| \sum_{n=l+1}^{k} T^n g \|_p &\leq \lim_{k,l\to\infty} \sum_{n=l+1}^{k} \| T^n g \|_p \\ &\leq \lim_{k,l\to\infty} \sum_{n=l+1}^{k} \| T^n \|_p \| g \|_p \\ &\leq \| g \|_p \lim_{k,l\to\infty} \sum_{n=l+1}^{k} \| T \|_p^n, \end{aligned} \tag{31}$$

where $1 \leq p < 3/2$, as required in the present context. From Eq. (31) we infer that the Cauchy condition, Eq. (30), is satisfied if

$$\| T \|_p < 1, \tag{32}$$

which thus represents the condition sufficient for the Neumann series, Eq. (29), to converge in the mean of index $1 \leq p < 3/2$ to a function $f^*(\bar{x}) \in L_p(D)$.

This sufficient condition is satisfied *a fortiori* for any $1 \leq p < 3/2$ when

$$A = \sum_{\mathrm{sf}}^{\max} \frac{1 - e^{-\Sigma_{\min}\Delta}}{\Sigma_{\min}} < 1 \tag{33a}$$

for $\Sigma_{\min} > 0$, or when

$$A = \sum_{\mathrm{sf}}^{\max} \Delta < 1 \tag{33b}$$

for $\Sigma_{\min} = 0$.

Equations (33) relate thus the convergence of the Neumann series to the nuclear and geometrical characteristics of the body under examination.

That $f^*(\bar{x})$, Eq. (29), is a solution of Eq. (1) follows from the continuity of the transformation T. Apart from a set of zero measure it is also the unique solution of Eq. (1). In fact, if $f^{**}(\bar{x})$ were another solution of Eq. (1), one may write

$$f^* - f^{**} = T(f^* - f^{**}) \tag{34a}$$

and, therefore

$$(1 - \| T \|_p) \| f^* - f^{**} \|_p \leq 0, \tag{34b}$$

which implies

$$\| f^* - f^{**} \|_p = 0, \tag{34c}$$

according to Eq. (32).

The results quoted above for the existence and for the uniqueness of the solution of Eq. (1) via Neumann series representation hold also when the free term $g(\bar{x})$, Eq. (25), is generalized as

$$g(\bar{x}) = \sum_{n=1}^{N} Q_n \frac{e^{-\tau(\bar{x},\bar{x}_n)}}{4\pi |\bar{x} - \bar{x}_n|^2}, \tag{35}$$

that is, when a finite number N of isotropic delta-like sources—of intensity $Q_1, \ldots, Q_n$—are localized at the points $\bar{x}_1, \ldots, \bar{x}_n$ in the interior of D.

The specialization to the case of a homogeneous body, that is, when cross sections are independent of position, is also immediate. In this case the kernel $k(\bar{x}, \bar{y})$, Eq. (24), becomes simply

$$k(\bar{x}, \bar{y}) = \sum_{\mathrm{sf}} \frac{e^{-\Sigma|\bar{x}-\bar{y}|}}{4\pi |\bar{x} - \bar{y}|^2}, \tag{36}$$

which is a self-adjoint kernel with weak singularity, according to the definition of Eq. (22). It is not a kernel of convolution type like in Eq. (21a), as the domain D is not coincident with R_3.

The free term $g(\bar{x})$, Eq. (25), becomes instead

$$g(\bar{x}) = Q_1 \frac{e^{-\Sigma|\bar{x}-\bar{x}_1|}}{4\pi |\bar{x} - \bar{x}_1|^2}. \tag{37}$$

As a comment on the application of Theorem 1 we underline the following two points:

(i) For the case of the homogeneous body, that is, for the problem represented by Eqs. (1), (36) and (37), an alternative to the present method is the one proposed in a previous paper,[3] where the domain D is specialized as a sphere, cylinder or parallelepiped, and consists of subjecting both sides of Eq. (1) to a three-dimensional infinite Fourier transform of vector $\bar{B}$. To do this the original domain D is first mapped into the three-dimensional Euclidean space R_3 by supposing that the body is surrounded by a purely absorbing material occupying the residual infinite space $R_3 - D$. The macroscopic absorption cross section Σ_a^* of this external material is taken to be equal to the total macroscopic cross section Σ of the body.

The details of this procedure can be found in Ref. 3 (for a similar approach see also Refs. 4 and 5). This procedure amounts just to a sort of regularization of Eq. (1), as it can be verified that the Fourier transform $\tilde{g}(\bar{B})$ of $g(\bar{x})$, Eq. (37), and the kernel $\tilde{k}(\bar{B}, \bar{B}')$ of the resulting transformed equation for the transform $\tilde{f}(\bar{B})$ of the solution $f(\bar{x})$ of Eq. (1) belong to $L_p(R_3)$ and $N_p(R_3)$ with the same index $p > 3$, respectively.

Thus the theory of linear integral equations with kernel of finite double norm can be exploited for proving the existence and the uniqueness of $\tilde{f}(\bar{B})$ via Neumann series representation. Then one is left with the task of casting $\tilde{f}(\bar{B})$ in a form suitable for the final inversion.[3]

(ii) For the case when the source is not any longer an isotropic delta-like source, but it is distributed throughout the inhomogeneous body and is represented by a bounded real nonnegative measurable function $Q(\bar{x})$ for $\bar{x} \in D$, the free term $g(\bar{x})$, Eq. (25), becomes

$$g(\bar{x}) = \int_D Q(\bar{y}) \frac{e^{-\tau(\bar{x},\bar{y})}}{4\pi |\bar{x} - \bar{y}|^2} d\bar{y}, \tag{38}$$

which is bounded on D and belongs to any Lebesgue space $L_p(D)$ with $p \geq 1$.

In this case the problem [Eqs. (1), (24) and (38)] can again be handled in the frame of the usual theory of linear integral equations with kernel of finite double norm. There results that the conditions sufficient for

the Neumann series to be convergent are just the same as the ones already expressed by Eqs. (33a) and (33b), and the solution $f(\bar{x})$ of Eq. (1) is bounded on D.[3] Moreover, the solution can be given a practical representation by resorting to a suitable polynomial degeneration of the kernel $k(\bar{x},\bar{y})$, Eq. (24).

It is then understood why the problem, Eqs. (1), (24) and (25), requires a treatment *ad hoc*, as worked out in the present investigation. In fact it is at once realized that:

A Fourier transform regularization like in the homogeneous case (i) cannot be performed because of the functional structure the kernel $k(\bar{x},\bar{y})$, Eq. (24), assumes as a consequence of the nuclear heterogeneity of the body;

As will be shown later on, the solution $f(\bar{x})$ of Eq. (1) is not any longer bounded on D, as in case (ii) of a distributed source, because of the singularity of the free term $g(\bar{x})$, Eq. (25). We know only that $f(\bar{x})$ is $L_p(D)$ with $1 \leq p < 3/2$. This latter circumstance prevents for now constructing a practical solution of Eq. (1) via kernel degeneration.

We conclude by turning to Theorem 2. We see that $k(\bar{x},\bar{y})$, Eq. (24), satisfies conditions, Eqs. (13a) and (13b), with

$$1 \leq p' < 3/2 \tag{39a}$$

as follows from Eq. (23c). Then

$$4/3 < 1/p + 1/p' \leq 2 \tag{39b}$$

and thus, from Eq. (15),

$$1 \leq r < 3. \tag{39c}$$

Hence Tg, where g is $L_p(D)$ with $1 \leq p < 3/2$, belongs to $L_r(D)$, with r given by Eq. (39c).

By repeated application of Theorem 2 it is easily inferred that:

T^2g is $L_r(D)$, with $1 \leq r < \infty$;

T^3g is bounded and therefore $L_r(D)$, with $1 \leq r \leq \infty$;

$T^ng, n \geq 4$ is bounded and $L_r(D)$, with $1 \leq r \leq \infty$, like T^3g and furthermore continuous on D, as the kernel, Eq. (24), is completely continuous in the Mikhlin sense, if $\sum(\bar{y})$ and $\sum_{sf}(\bar{y})$ exhibit at most a finite number of discontinuity surfaces.[3]

If we recall that the general term T^ng of the Neumann series represents the flux of neutrons which have undergone n collisions in the interior of the body before reaching the point $\bar{x}$, it is clear that the effects of repeated application of the transformation T to the virgin flux $g(\bar{x})$ amount—as physically expected—to equalize the neutron distribution and to smooth out rapid fluctuations starting from the singularity $g(\bar{x})$ shows at the point at which the source is localized.

The Neumann series is thus verified to converge in the mean of index r to a function $f(\bar{x}) \in L_r(D)$, with at most $1 \leq r < 3/2$, which is the unique solution of Eq. (1), whenever Eq. (33a) or Eq. (33b) is satisfied.

By the same procedure the series

$$h(\bar{x}) = \sum_{n=4}^{\infty} T^ng \tag{40}$$

is seen to converge uniformly to a bounded continuous function, as $\|T\|_\infty < 1$ according to Eqs. (33) and (28). The solution $f(\bar{x})$ of Eq. (1) is then the superposition of $h(\bar{x})$, Eq. (40), and of the function $g + Tg + T^2g + T^3g$, which belongs only to a Lebesgue space $L_p(D)$, with $1 \leq p < 3/2$, and is not bounded on D.

The solution $f(\bar{x})$ of Eq. (1) is thus no longer continuous and bounded on D.

[1] A. C. Zaanen, *Linear Analysis*, (North-Holland, Amsterdam, 1964).
[2] U. Neri, "Singular Integrals" in *Lecture Notes in Mathematics* (Springer-Verlag, Berlin, 1971).
[3] V. C. Boffi, F. Premuda, and G. Spiga, Math. Phys. 14, 346 (1973).
[4] K. M. Case and R. D. Hazeltine, Math. Phys. 11, 1126 (1970).
[5] K. M. Case and R. D. Hazeltine, Math. Phys. 12, 1970 (1971).

Infinite spin dimensionality limit for nontranslationally invariant interactions

Hubert J. F. Knops*

Baker Laboratory, Cornell University, Ithaca, New York 14850
(Received 25 May 1973)

A proof is presented that a nontraslationally invariant system of classical n-dimensional spins approaches in the limit $n \to \infty$ a suitable generalized spherical model.

1. INTRODUCTION

It has been observed[1] that the spherical model[2] can be considered as a limit of infinite spin dimensionality ($n \to \infty$). A rigorous proof of this fact was recently given by Kac and Thomson,[3] who showed, moreover, that the result was independent of the order in which the thermodynamical limit (number of spins, $N \to \infty$) and the infinite spin dimensionality limit are taken. Their analysis is restricted, however, to translationally invariant interactions.

In the present note it is shown that an analogous theorem can be proven for nontranslation invariant interactions. It turns out that the limiting model in this case is not the usual spherical model but is, rather, a more general model with multiple spherical constraints. In the case of periodic interaction this model is similar to the so-called "m-spherical model" discussed by Mazo.[4] The proof that will be given is essentially an adaptation of the proof of Kac and Thomson to the present case, parts of their proof have, however, been simplified.

Situations in which the present result may be of interest include the study of finite size effects and impurities. Recent results[5] indicate that there is a discrepancy between the shifts in critical temperature due to a finite size when calculated respectively for "fixed spin" models or for the spherical model. The present result suggests that this discrepancy might well disappear if the shifts were calculated for the generalized spherical model considered here.

Another instance in which the present results apply is that of an antiferromagnet in a staggered magnetic field. The properties of the generalized spherical model obtained in this case have been explicitly calculated.[6]

2. INFINITE SPIN DIMENSIONALITY LIMIT

Consider a lattice of N spins, interacting through a coupling $-\rho_{i,j}(N)$. Assume that the eigenvalues of these matrices are uniformly bounded in N. In the case of a nontranslation invariant interaction we extend the notion of a spherical model in the following way. Define a partition function Q_N by

$$Q_N(\beta; z_1, \ldots, z_N) \equiv \int_{-\infty}^{\infty} \cdots \int \exp\left(\sum_{i,j} (\beta\rho_{i,j}(N) - \delta_{i,j} z_i) S_i S_j\right) dS_1 \cdots dS_N. \quad (2.1)$$

This definition applies, of course, only in a region D where the "spherical fields" z_j are large enough to ensure that the quadratic form in the exponent is strictly negative definite. The spherical fields are then to be determined by the individual spherical constraints, which are

$$1 = \langle S_j^2 \rangle = -(\partial \ln Q_N / \partial z_j), \quad j = 1, \ldots, N. \quad (2.2)$$

The following argument shows that these equations indeed determine the spherical fields uniquely as a function of β and N. Consider the function

$$F_N(\beta; z_1, \ldots, z_N) \equiv \ln Q_N + \sum_j z_j. \quad (2.3)$$

Some reflection shows that this function approaches $+\infty$ everywhere at the boundary of its definition region D. Moreover, the function is jointly convex in the z_j and everywhere differentiable with nonconstant derivatives. Consequently, it has a unique minimum which is just determined by the relations (2.2).

Upon denoting the solutions of (2.2) by $z_j^N(\beta)$, we may define the limiting "spherical" free energy density by

$$f(\beta) = \lim_{N\to\infty} N^{-1} F_N(\beta; z_1^N(\beta), \ldots, z_N^N(\beta)). \quad (2.4)$$

In view of the discussion above, one can equivalently write

$$f(\beta) = \lim_{N\to\infty} N^{-1} \min_{\{z_1, \ldots, z_N\} \in D} F_N(\beta; z_1, \ldots, z_N). \quad (2.5)$$

We stress that the minimum is to be taken before the limit $N \to \infty$.

In the case of a periodic interaction it is clear from the uniqueness that it suffices to satisfy (2.2) for the spins of a unit cell, hence reducing the number of independent spherical fields drastically. One can then use the periodicity of the interactions to obtain a partial diagonalization of the quadratic form. This leads (for the periodic case) to a proof of the existence of the limits (2.4) or (2.5) showing, in fact, that the processes of taking the limit and finding the minimum may be interchanged in (2.5). The proof rests essentially on the observation that the conditions (2.2) prevent the spherical fields from approaching the boundary of the region D too fast in the limit $N \to \infty$. We shall omit the details of the proof since it is not directly relevant for the rest of our discussion.

Notice further that, for a periodic interaction, one can also write (2.2) equivalently as

$$\left\langle \sum_{j\in P} S_j^2 \right\rangle = N_P, \quad \text{for all } P, \quad (2.6)$$

in which the sums are to be taken over sublattices P of sites equivalent under the periodicity. Evidently, N_P is the number of those sites in a sublattice. In this way the spherical constraints involve averages of macroscopic quantities. It can thus be demonstrated[2-4] that in the thermodynamic limit one can equally well replace these average constraints by the strict constraints

$$\sum_{j\in P} S_j^2 = N_P, \quad \text{for all } P. \quad (2.7)$$

These remarks apply also for the systems of finite

thickness, treated by Fisher and Barber,[5] since these can be considered as a special case of periodicity.

The theorem we want to prove can now be formulated.

Theorem: Let the partition function of a system of n-dimensional spins of length $n^{1/2}$ be defined by

$$Z_N^n(\beta) \equiv \int\cdots\int_{|\mathbf{S}_j|=n^{1/2}} \exp\left(\sum_{i,j} \beta\rho_{i,j}(N)\mathbf{S}_i\cdot\mathbf{S}_j\right) d\mathbf{S}_1\cdots d\mathbf{S}_N, \quad (2.8)$$

then, with $f(\beta)$ as in (2.4), one has

$$\lim_{N,n\to\infty} (Nn)^{-1}\ln Z_N^n(\beta) = f(\beta), \quad (2.9)$$

whenever the limit (2.4) defining the spherical free energy exists. This is true independent of the order in which N and n tend to infinity.

To prove this result, notice first that the addition of a constant times the unit matrix to the interaction merely results in the subtraction of β times this constant from both sides of (2.9). In view of the lower bound that we assumed on $\rho_{i,j}(N)$ one can always choose this constant so that the resulting matrices become positive definite. In the following we can, without loss of generality, hence assume that $\rho_{i,j}(N)$ is positive definite.

It is useful to define a more general form of the partition function for a system of n-dimensional spins by

$$Z_N^n(\beta;\lambda_1,\ldots,\lambda_N) \equiv \int\cdots\int_{|\mathbf{S}_j|=n^{1/2}\lambda_j} \times \exp\left(\sum_{i,j}\beta\rho_{i,j}(N)\mathbf{S}_i\cdot\mathbf{S}_j\right) d\mathbf{S}_1\cdots d\mathbf{S}_N. \quad (2.10)$$

Direct calculation yields

$$\int_0^\infty\cdots\int\left[Z_N^n(\beta;\lambda_1,\ldots,\lambda_N)\exp\left(-\sum_j nz_j\lambda_j^2\right)\right]d\lambda_1\cdots d\lambda_N = n^{-N/2}Q_N(\beta;z_1,\ldots,z_N)^n. \quad (2.11)$$

By a trivial regrouping of terms this can also be written as

$$\int_0^\infty\cdots\int\left[Z_N^n(\beta;\lambda_1,\ldots,\lambda_N)\exp\left(-\sum_j nz_j'\lambda_j^2\right)\right] \times \exp\left(-(\pi/2)\sum_j\lambda_j^2\right)d\lambda_1\cdots d\lambda_N = n^{-N/2}Q_N(\beta;z_1,\ldots,z_N)^n, \quad (2.12)$$

in which $z_j' = z_j - (\pi/2n)$.

When the term inside the braces is replaced by its maximum the remaining integral is unity and one concludes

$$n^{-N/2}Q_N(\beta;z_1,\ldots,z_N)^n \le \max_{0\le\lambda_1,\ldots,\lambda_N<\infty}\left[Z_N^n(\beta;\lambda_1,\ldots,\lambda_N) \times \exp\left(-n\sum_j z_j'\lambda_j^2\right)\right]. \quad (2.13)$$

Upon taking logarithms one obtains

$$\max_{0\le\lambda_1,\ldots,\lambda_N<\infty}\left((Nn)^{-1}\ln Z_N^n(\beta;\lambda_1,\ldots,\lambda_N) - N^{-1}\sum_j z_j\lambda_j^2 + \pi(2nN)^{-1}\sum_j\lambda_j^2 + (2n)^{-1}\ln n\right) \ge N^{-1}\ln Q_N(\beta;z_1,\ldots,z_N). \quad (2.14)$$

By adjusting $z_1,\ldots,z_N$ the maximum can be made to occur at any set of positive values of $\lambda_1,\ldots,\lambda_N$. Specifically one can find values $z_1^*,\ldots,z_N^*$ for the spherical fields such that the maximum occurs at the point $\lambda_1 = 1,\ldots,\lambda_N = 1$. The value of the maximum at this point is a finite number; consequently, $Q_N(\beta;z_1^*,\ldots,z_N^*)$ must be finite and hence $z_1^*,\ldots,z_N^* \in D$. For these special values the inequality takes the form

$$(Nn)^{-1}\ln Z_N^n(\beta) + \epsilon_n \ge N^{-1}\ln Q_N(\beta;z_1^*,\ldots,z_N^*) + N^{-1}\sum_j z_j^*, \quad (2.15)$$

in which

$$\epsilon_n = (2n)^{-1}(\pi + \ln n). \quad (2.16)$$

The result of taking the minimum of all allowed z_j values on the rhs can only be the sharpening of this inequality. Together with the definition (2.3) this yields

$$(Nn)^{-1}\ln Z_N^n(\beta) + \epsilon_n \ge \min_{\{z_1,\ldots,z_N\}\in D} N^{-1}F_N(\beta;z_1,\ldots,z_N). \quad (2.17)$$

Notice that, by taking n large, ϵ_n can be made arbitrarily small uniformly in N.

The next step is to prove a similar but reversed inequality. In order to do this, consider the well-known identity

$$\exp\left(\beta\sum_{i,j}\rho_{i,j}(N)\mathbf{S}_i\cdot\mathbf{S}_j\right) = [\det\rho_{i,j}(N)]^{-n/2}(2\pi)^{-Nn/2}\int_{-\infty}^{\infty}\cdots\int d\mathbf{x}_1\cdots d\mathbf{x}_N \times \exp\left(-\tfrac{1}{2}\sum_{i,j}[\rho_{i,j}^{-1}(N)\mathbf{x}_i\cdot\mathbf{x}_j] + (\sqrt{2\beta})\sum_i\mathbf{S}_i\cdot\mathbf{x}_i\right), \quad (2.18)$$

which is valid for any positive definite $\rho(N)$ and n-dimensional vectors $\mathbf{S}_j$. With the help of this relation one can rewrite the formula (2.8), defining Z_N^n, in the form

$$Z_N^n(\beta) = [\det\rho_{i,j}(N)]^{-n/2}(2\pi)^{-Nn/2}\int_{-\infty}^{\infty}\cdots\int d\mathbf{x}_1\cdots d\mathbf{x}_N \times \exp[-\tfrac{1}{2}\sum_{i,j}(\rho_{i,j}^{-1}(N) - \delta_{i,j}\beta/z_j)\mathbf{x}_i\cdot\mathbf{x}_j]\prod_j\Phi_j((\sqrt{2\beta})\mathbf{x}_j), \quad (2.19)$$

in which

$$\Phi_j(\mathbf{y}) \equiv \int_{|\mathbf{S}|=n^{1/2}}\exp(\mathbf{S}\cdot\mathbf{y})d\mathbf{S}\cdot\exp(-y^2/4z_j). \quad (2.20)$$

The diagonal terms which have been added to $\rho^{-1}(N)$ are compensated by the exponential factor in the definition of Φ_j. By passing to polar coordinates the integral defining Φ_j can be written

$$A_n\int_0^\pi \exp(n^{1/2}y\cos\theta)(\sin^2\theta)^{(n/2)-1}d\theta, \quad (2.21)$$

where A_n is the surface area of a $(n-1)$-dimensional sphere given by

$$A_n = 2(\pi n)^{(n-1)/2}/\Gamma(\tfrac{1}{2}n - \tfrac{1}{2}). \quad (2.22)$$

Replacement of the integrand in (2.21) by its maximum, which occurs at a point θ_0 given by

$$n^{1/2}y\cos\theta_0 = (n/2-1)\{-1 + [1 + 4ny^2/(n-2)^2]^{1/2}\}, \quad (2.23)$$

yields the upper bound

$$\Phi_j(y) \le A_n\exp[(n/2-1)(-1 + \ln 2 + [1 + 4ny^2/(n-2)^2]^{1/2} - \ln\{1 + [1 + 4ny^2/(n-2)^2]^{1/2}\}) - y^2/4z_j]. \quad (2.24)$$

Calculation of the maximum over all y of this bound readily yields an overall majorization on Φ_j, namely

$$\Phi_j \leq A_n \exp[(n/2)(2z_j - 1 - \ln 2z_j + \epsilon_j')], \quad (2.25)$$

where

$$\epsilon_j' = n^{-1}(2 + 2\ln 2z_j). \quad (2.26)$$

Replacement of Φ_j by this bound in (2.19) reduces the integral to a simple Gaussian integral. Some reflection shows that the quadratic form $\rho^{-1}_{i,j}(N) - \delta_{i,j}\beta/z_i$ is positive definite whenever $\beta\rho_{i,j}(N) - \delta_{i,j}z_j$ is negative definite. Consequently, the integration can be carried out for all $z_j \in D$, yielding the upper bound

$$Z_N^n(\beta) \leq [\det\rho_{i,j}(N)]^{-n/2}\{\det[\rho_{i,j}(N)^{-1} - \delta_{i,j}\beta/z_j]\}^{-n/2} \times (A_n)^N \exp\left((n/2)\sum_j (2z_j - 1 - \ln 2z_j + \epsilon_j')\right). \quad (2.27)$$

Taking logarithms and multiplying the determinants yields

$$(Nn)^{-1}\ln Z_N^n(\beta) \leq -(2N)^{-1}\ln\{\det[z_j\delta_{i,j} - \beta\rho_{i,j}(N)]\} + N^{-1}\sum_j z_j - \tfrac{1}{2} - \tfrac{1}{2}\ln 2 + n^{-1}\ln A_n + (2N)^{-1}\sum_j \epsilon_j'. \quad (2.28)$$

We may use Stirling's formula for A_n to obtain for large n the result

$$(Nn)^{-1}\ln Z_N^n(\beta) \leq -(2N)^{-1}\ln\{\det[z_j\delta_{i,j} - \beta\rho_{i,j}(N)]\} + \tfrac{1}{2}\ln\pi + N^{-1}\sum_j z_j + (2N)^{-1}\sum_j \epsilon_j'. \quad (2.29)$$

Calculation of the Gaussian integral defining $Q_N(\beta; z_1, \ldots, z_N)$ in (2.1) reveals that the logarithm of this integral is given by the first two terms of the rhs of the inequality above. On further recalling the definition (2.3) of F_N, the inequality takes the form

$$(Nn)^{-1}\ln Z_N^n(\beta) \leq N^{-1}F_N(\beta; z_1, \ldots, z_N) + (2N)^{-1}\sum_j \epsilon_j'. \quad (2.30)$$

Since this inequality applies for all $z_j \in D$, one also has

$$(Nn)^{-1}\ln Z_N^n(\beta) \leq N^{-1}\min_{\{z_1,\ldots,z_N\}\in D} F_N(\beta; z_1, \ldots, z_N) + (2N)^{-1}\sum_j \epsilon_j'. \quad (2.31)$$

Consider now the expression (2.26) for the errors ϵ_j' calculated for the values of z_j minimizing F_N, i.e., $z_j = z_j^N(\beta)$. From the fact that $\rho_{i,j}(N)$ is positive definite one easily derives the following lower bounds on F_N:

$$N^{-1}F_N(\beta; z_1^N(\beta), \ldots, z_N^N(\beta)) \geq N^{-1}\sum_j [z_j^N(\beta) - \tfrac{1}{2}\ln z_j^N(\beta)] > N^{-1}\sum_j \ln z_j^N(\beta). \quad (2.32)$$

The lhs of these inequalities converges by assumption when $N \to \infty$. Consequently, the rhs has to be bounded in N. It follows then that also $(2N)^{-1}\sum_j \epsilon_j'$ is bounded in N and that in fact, since each term contains a factor n^{-1}, this sum can be made arbitrarily small for large n, uniformly in N.

Combination of the inequality (2.31) with its reversed form (2.17) proves then the theorem.

ACKNOWLEDGMENTS

It is a pleasure to thank Professor Michael Fisher for suggesting the problem and a critical reading of the manuscript. The award of a fellowship by the Niels Stensen Stichting and the auxiliary support of the National Science Foundation partly through the Materials Science Center at Cornell University is gratefully acknowledged.

* On leave of absence from University of Nymegen, Nymegen, Netherlands.

[1] H. E. Stanley, Phys. Rev. 176, 718 (1968).
[2] T. H. Berlin and M. Kac, Phys. Rev. 86, 821 (1952).
[3] M. Kac and C. J. Thomson, Physica Novegica 5, 3 (1971).
[4] G. Bettoney and R. M. Mazo, J. Math. Phys. 11, 1147 (1970).
[5] M. E. Fisher and M. N. Barber, Phys. Rev. Lett. 28, 1516 (1972); Ann. Phys. (N.Y.) (1973).
[6] H. J. F. Knops, Phys. Rev., to be published.

Probability density function and moments of the field in a slab of one-dimensional random medium

R. H. Lang

Department of Electrical Engineering and Computer Science, George Washington University, Washington, D.C.
(Received 15 December 1972)

The problem of a plane wave normally incident on a slab of one-dimensional random medium is studied. The refractive index variations of the random medium are taken to be a stationary Gaussian–Markov process. By employing an invariant imbedding technique and by using the Markov property of the refractive index variations, two cascaded diffusion equations are obtained for the probability density function of the reflection coefficient and the field in the slab. These equations are then solved approximately for small refractive index fluctuations and an expression is obtained for mean intensity in the slab interior.

1. INTRODUCTION

This paper is a study of wave propagation in a slab of one-dimensional random medium. The slab has width $\bar{L}$; the medium inside the slab has a refractive index variation $n(\bar{x}) = [1 + \epsilon\xi(\bar{x})]^{1/2}$, $0 \le \bar{x} \le \bar{L}$. We assume that $\xi(\bar{x})$ is a stationary Gaussian-Markov process with zero mean and exponential correlation[1] (Ornstein-Uhlenbeck process) and that ϵ is a small parameter. The regions $\bar{x} > \bar{L}$ and $\bar{x} < 0$ are homogeneous regions with unit refractive indexes; and a plane wave is normally incident on the slab from the region $\bar{x} > \bar{L}$.

This problem has been the subject of a number of investigations which have appeared in the literature. The investigators have directed their efforts toward finding approximate expressions for the moments of various statistical quantities associated with the slab such as reflection coefficient, transmission coefficient, and the field. Gertsenshtein and Vasil'iev[2] considered a medium composed of discrete random inhomogeneities. They found an expression for the mean square reflection coefficient in the limit as their discrete medium approached a continuum. Gazaryan,[3] again using a discrete model, found expressions for the mean field and intensity of the field in the continuum limit. The problem was then treated by a number of investigators who used a continuum model of the random medium directly instead of first considering a discretized medium. Kupiec, *et al.*, [4] found an expression for the mean field by applying the method of smooth perturbations to the Dyson equation. Papanicolaou[5] and Morrison, Papanicolaou, and Keller[6] have found expressions for the probability density function (p.d.f.) of the transmission coefficient and from this have calculated the mean squared transmission and reflection coefficients.

In this paper we employ the medium model used by Morrison-Papanicolaou-Keller and find the p.d.f. of the reflection coefficient and the field inside the slab. From these p.d.f.'s we obtain an expression for the mean intensity in the slab and, in addition, recover the results of Refs. 4, 5, and 6. Our expression for the mean intensity is similar to Gazaryan's, however, his result has been obtained by using different methods and a different model.

In Sec. 2 we start by reformulating the original problem as a boundary value problem over the interval $[0, \bar{L}]$. The boundary value problem is then imbedded in two cascaded initial value problems by employing an invariant imbedding technique.[7,8] The solution to the first initial value problem will be called the generalized reflection coefficient Γ. This solution provides initial conditions for the second initial value problem whose solution yields the desired field u.

The probabilistic nature of the problem is introduced by first representing Γ in terms of its amplitude and phase, i.e., $\Gamma = \rho \exp i\phi$ and then recognizing that $\mathbf{X} = (\rho, \phi, \xi)$ is a vector Markov process. As a result, a forward Kolmogorov equation can be written for the p.d.f. of $\mathbf{X}$. In a similar manner, one finds that the second initial value problem generates a vector Markov process $\mathbf{Y} = (v, \theta, \rho, \phi, \xi)$. Here v and θ are the log amplitude and phase of the field u, respectively. A second forward Kolmogorov equation can then be written for the p.d.f. of $\mathbf{Y}$. The initial distribution of the random variable $\mathbf{Y}$ is obtained from the solution of the first diffusion equation. Thus the original stochastic problem for the field u has been replaced by the deterministic problem of solving two cascaded diffusion equations.

In Sec. 3-5, we find approximate solutions to the diffusion equations for small ϵ. Our perturbation method follows the one that Morrison-Papanicolaou-Keller[6] used. Once approximate expressions for the p.d.f.'s of $\mathbf{X}$ and $\mathbf{Y}$ have been developed, the moments of the reflection coefficient and field are calculated.

2. FORMULATION

We consider a plane wave which is normally incident on a slab of one-dimensional random medium. The field u obeys the one-dimensional wave equation

$$\frac{d^2u}{d\bar{x}^2} + K^2 n^2(\bar{x})u = 0, \quad -\infty < \bar{x} < \infty, \tag{2.1}$$

where we require that $du/d\bar{x}$ be continuous. Here we take K to be the free space wavenumber and $n(\bar{x})$ to be

$$n(\bar{x}) = \begin{cases} 1, & \bar{x} < 0, \\ [1 + \epsilon\xi(\bar{x})]^{1/2}, & 0 \le \bar{x} \le \bar{L}, \\ 1, & \bar{x} > \bar{L}. \end{cases} \tag{2.2}$$

If we assume that the process $\xi(\bar{x})$ has a correlation length l and if we define

$$x = \bar{x}/l, \quad L = \bar{L}/l, \quad k = lK, \quad \mu(x) = \xi(\bar{x}), \tag{2.3}$$

then (2.1) can be put in the following normalized form:

$$\frac{d^2u}{dx^2} + k^2[1 + \epsilon\mu(x)]u = 0, \quad 0 \le x \le L,$$

in the slab region. Outside the slab, the normalized solution can be obtained explicitly. It is

$$u(x) = \begin{cases} e^{-ik(x-L)} + Re^{+ik(x-L)}, & x \ge L, \\ Te^{-ikx}, & x \le 0, \end{cases} \tag{2.5}$$

Here R and T are called the reflection and transmission coefficients of the slab.

The problem on the infinite interval can be reformulated in terms of the following problem on the finite interval:

$$\frac{d^2u(x)}{dx^2} + k^2[1 + \epsilon\mu(x)]u(x) = 0, \quad 0 \le x \le L, \tag{2.6a}$$

$$\frac{du(0)}{dx} + iku(0) = 0, \tag{2.6b}$$

$$\frac{du(L)}{dx} - iku(L) = -2ik. \tag{2.6c}$$

The boundary conditions have been obtained by using (2.5) along with the continuity of u and du/dx at $x = L$. We also find that

$$R = u(L) - 1, \quad T = u(0). \tag{2.7}$$

Next we convert the boundary value problem (2.6) to an initial value problem. This is accomplished by using an invariant imbedding procedure.[7,8] To apply this procedure, we exhibit the dependence of the field u on the slab width L explicitly, i.e., $u = u(x, L)$. Next we replace L by the variable t, $0 \le t \le L$, and note that $u(x, t)$ satisfies (2.6) with L replaced by t. Thus we have

$$\left(\frac{d^2}{dx^2} + k^2[1 + \epsilon\mu(x)]\right) u(x, t) = 0, \quad 0 \le x, t \le L, \tag{2.8a}$$

$$\frac{d}{dx} u(0, t) + iku(0, t) = 0, \tag{2.8b}$$

$$\frac{d}{dx} u(t, t) - iku(t, t) = -2ik. \tag{2.8c}$$

In Appendix A, we convert the above differential equation in x with t as a parameter to two cascaded initial value problems in t with x as a parameter. These initial value problems are:

Problem 1:

$$\frac{d\Gamma(t)}{dt} = 2ik\Gamma(t) + \tfrac{1}{2}i\epsilon k\mu(t)[1 + \Gamma(t)]^2, \tag{2.9a}$$

$$\Gamma(0) = 0, \quad 0 \le t \le x; \tag{2.9b}$$

Problem 2:

$$\frac{du(x, t)}{dt} = iku(x, t) + \tfrac{1}{2}i\epsilon k\mu(t)[1 + \Gamma(t)]u(x, t), \tag{2.10a}$$

$$u(x, x) = 1 + \Gamma(x), \tag{2.10b}$$

$$\frac{d\Gamma(t)}{dt} = 2ik\Gamma(t) + \tfrac{1}{2}i\epsilon k\mu(t)[1 + \Gamma(t)]^2, \tag{2.11a}$$

$$\Gamma(t)|_{t=x} = \Gamma(x), \quad x \le t \le L, \tag{2.11b}$$

The solution to the first problem obeys a Ricatti equation. We will call this solution, $\Gamma(t)$, the generalized reflection coefficient since $\Gamma(L) = R$ as is shown in (A7) of Appendix A. Once $\Gamma(t)$, $0 \le t \le x$, is determined, it provides the initial conditions for the second problem. The solution to the second problem provides us with $u(x, t)$, $x \le t \le L$, and thus we have the solution to the original boundary value problem (2.6) by setting $t = L$ in $u(x, t)$.

Since $\mu(t)$ is a Gaussian process, it can be specified completely by its first two moments. They are

$$\langle\mu(t)\rangle = 0, \quad \langle\mu(t)\mu(t + \tau)\rangle = e^{-|\tau|}, \tag{2.12}$$

where the brackets indicate the expected values of the quantities enclosed. It is also noted that the process $\mu(t)$ can be generated from the Itô stochastic differential equation[9,10]

$$d\mu = -\mu dt + \sqrt{2}\, d\beta, \tag{2.13}$$

where $\beta(t)$ is the standard Brownian motion process. The distribution of μ at $t = 0$ is taken to be $(2\pi)^{-1/2} \times \exp(-\mu^2/2)$.

Equations (2.13) and (2.9) can be solved simultaneously. Before doing this, however, we rewrite Γ in terms of its amplitude and phase, i.e.,

$$\Gamma = \rho e^{i\phi}, \quad 0 \le \rho \le 1, \quad -\pi < \phi \le \pi. \tag{2.14}$$

Upon substituting (2.14) into (2.9a) and equating real and imaginary parts to zero, we obtain two equations involving ρ and ϕ. When these two equations are considered along with (2.13), we have the following system of stochastic differential equations:

$$d\begin{pmatrix}\rho\\ \phi\\ \mu\end{pmatrix} = \begin{pmatrix}2^{-1}k\epsilon\mu(1-\rho^2)\sin\phi\\ 2k + 2^{-1}k\epsilon\mu[2 + (\rho + \rho^{-1})\cos\phi]\\ -\mu\end{pmatrix} dt + \sqrt{2}\, d\begin{pmatrix}0\\ 0\\ \beta\end{pmatrix}, \quad 0 \le t \le x. \tag{2.15}$$

The initial conditions for the above system are: $\rho = 0$ with probability one; ϕ is distributed uniformly;[11] and u has the initial distribution associated with (2.13). Noting that the initial data is independent of the Brownian increments, $d\beta$, $0 \le t \le x$, one can show that (2.15) generates a three-dimensional Markov process, $\mathbf{X} = (\rho, \phi, \mu)$ and that the p.d.f. $p_1(\mathbf{X}, t)$ obeys the forward Kolmogorov equation[10]

$$\frac{\partial p_1}{\partial t} = L_1 p_1, \quad L_1 = L_1^{(0)} + \epsilon L_1^{(1)}, \tag{2.16}$$

where

$$L_1^{(0)} = -2k\frac{\partial}{\partial\phi} + \frac{\partial}{\partial\mu}\mu + \frac{\partial^2}{\partial\mu^2} \tag{2.17}$$

and

$$L_1^{(1)} = \frac{k\mu}{2}\{(\rho^2 - 1)\sin\phi\frac{\partial}{\partial\rho} - [2 + (\rho + \rho^{-1})\cos\phi]\frac{\partial}{\partial\phi} + (3\rho + \rho^{-1})\sin\phi\}. \tag{2.18}$$

The initial data for (2.16) is given by

$$p_1(\mathbf{X}, 0) = (2\pi)^{-3/2}\delta(\rho)e^{-\mu^2/2}, \tag{2.19}$$

where $\delta(\rho)$ is the Dirac delta function. By using p_1 evaluated at $t = L$ one can calculate the moments of R, but to obtain the moments of $u(x, L)$, one must consider the diffusion equation associated with Problem 2.

With this aim in mind, we introduce (2.14) and

$$u = e^{ik(t-x)e^{v+i\theta}}, \quad -\infty < v < \infty, \quad -\pi < \theta \leq \pi, \tag{2.20}$$

into (2.10) and (2.11). This results in four real equations for v, θ, ρ, and ϕ to which we append (2.13). The resulting system of stochastic differential equations is

$$d\begin{pmatrix} v \\ \theta \\ \rho \\ \phi \\ \mu \end{pmatrix} = \begin{pmatrix} -2^{-1}k\epsilon\mu\rho \sin\phi \\ 2^{-1}k\epsilon\mu(1+\rho\cos\phi) \\ 2^{-1}k\epsilon\mu(1-\rho^2)\sin\phi \\ 2k + 2^{-1}k\epsilon\mu[2+(\rho+\rho^{-1})\cos\phi] \\ -\mu \end{pmatrix} dt + \sqrt{2}\, d\begin{pmatrix} 0 \\ 0 \\ 0 \\ 0 \\ \beta \end{pmatrix}, \quad x \leq t \leq L. \tag{2.21}$$

The solution to (2.21) generates a five-dimensional Markov process, $\mathbf{Y} = (v, \theta, \rho, \phi, \mu)$, since the initial distribution of $\mathbf{Y}$ at $t = x$ is independent of $d\beta$, $x \leq t \leq L$. Therefore the p.d.f. $p_2(\mathbf{Y}, x, t)$ satisfies the following forward diffusion equation:

$$\frac{\partial p_2}{\partial t} = L_2 p_2, \quad L_2 = L_2^{(0)} + \epsilon L_2^{(1)}, \tag{2.22}$$

where $L_2^{(0)} = L_1^{(0)}$, which is defined in (2.17), and where

$$L_2^{(1)} = L_1^{(1)} + \frac{k\mu}{2}\left(\rho \sin\phi \frac{\partial}{\partial v} - (1+\rho\cos\phi)\frac{\partial}{\partial\theta}\right) \tag{2.23}$$

with $L_1^{(1)}$ being given by (2.18).

The initial distribution of $\mathbf{Y}$ at $t = x$ is obtained by using $p_1(\mathbf{X}, x)$ along with (2.10b) and (2.20). The result is

$$p_2(\mathbf{Y}, x, x) = p_1(\mathbf{X}, x)\delta[v - f(\rho, \phi)]\delta[\theta - g(\rho, \phi)], \tag{2.24}$$

where

$$f(\rho, \phi) = \tfrac{1}{2}\ln[1 + \rho^2 + 2\rho\cos\phi], \tag{2.25}$$

$$g(\rho, \phi) = \tan^{-1}[\rho \sin\phi/(1 + \rho\cos\phi)]. \tag{2.26}$$

3. PERTURBATION ANALYSIS OF PROBLEM 1

We will now find solutions to (2.16) for small ϵ. We start by representing $p_1(\mathbf{X}, t)$ in terms of the eigenfunctions of the operator L_1, i.e.,

$$p_1(\mathbf{X}, t) = \sum_q a_q e^{\lambda_q t} V_q(X), \tag{3.1}$$

where the eigenfunctions satisfy the equation

$$L_1 V_q = \lambda_q V_q. \tag{3.2}$$

Since L_1 is a three-dimensional operator, it should be noted that index q is also three dimensional.

Because the eigenfunctions V_q are difficult to obtain, we take advantage of the fact that L_1 is composed of a sum of an operator $L_1^{(0)}$, whose eigenfunctions are readily obtainable, plus a small operator $\epsilon L_1^{(1)}$. Expanding

$$V_q = \sum_n V_q^{(n)}\epsilon^n, \quad \lambda_q = \sum_n \lambda_q^{(n)}\epsilon^n \tag{3.3}$$

in a power series in ϵ, then plugging (3.3) into (3.2) and equating equal powers of ϵ, one obtains an infinite set of perturbation equations. The solution to the first of these equations yields $V_q^{(0)}$ and its corresponding eigenvalues $\lambda_q^{(0)}$. Higher order $V_q^{(n)}$ are then found by successively solving the higher order perturbation equations. The $\lambda_q^{(n)}$, $n = 1, 2, \cdots$, are obtained by requiring that the $V_q^{(n)}$, $n = 1, 2, \cdots$, by periodic in ϕ over $[0, 2\pi]$. This procedure for determining the $\lambda_q^{(n)}$ eliminates secular terms in t from occurring in the perturbation expansion of $p_1(X, t)$.

By employing these approximate eigenfunctions and eigenvalues in (3.1), we have

$$p_1(\mathbf{X}, t) = \sum_q a_q e^{(\lambda_q^{(0)} + \epsilon\lambda_q^{(1)} + \epsilon^2\lambda_q^{(2)})t} V_q^{(0)}(X) + O(\epsilon), \quad 0 \leq \epsilon^3 t < o(1). \tag{3.4}$$

Since the calculation of $V_q^{(0)}$ and $\lambda_q^{(n)}$, $n = 0, 1, 2$, is similar to the calculation that has been performed in Ref. 6, we will not include it here. Once these quantities have been found the a_q in (3.4) can be computed by using the initial data given by (2.19) and the orthogonality properties of the eigenfunctions. Next we integrate out the μ dependence from $p_1(\mathbf{X}, t)$ since it will not be required. We finally obtained

$$\hat{p}_1(\hat{\mathbf{X}}, t) = \int_{-\infty}^{+\infty} p_1(\mathbf{X}, t)\, d\mu = (2\pi)^{-1}(z-1)^{1/2}(z+1)^{3/2} \times \int_0^\infty s \tanh\pi s P_{-1/2+is}(z) e^{\lambda_0(s)\epsilon^2 t} ds + O(\epsilon^2), \quad 0 \leq \epsilon^3 t < o(1). \tag{3.5, 3.6}$$

where

$$z = (1+\rho^2)(1-\rho^2)^{-1},$$

$$\lambda_0(s) = -8^{-1}(1 + 4k^2)^{-1}k^2(4s^2 + 1) \tag{3.7}$$

with $\hat{\mathbf{X}} = (\rho, \phi)$. In the above $P_n(z)$ is the Legendre polynomial of order n. The fact that the expression for $\hat{p}_1$ is correct to $O(\epsilon^2)$ instead of $O(\epsilon)$ is because $\hat{p}_1$ is an even function of ϵ whereas p_1 is not.

The approximate p.d.f. given in (3.6) can now be used to find moments of the reflection coefficient by setting $t = L$. Because the approximate expression for $\hat{p}_1$ given by (3.6) is independent of ϕ, one immediately finds $\langle R^n \rangle = O(\epsilon^2)$, $n = 1, 2, \cdots$. In addition, the mean square reflection coefficient $\langle |R^2| \rangle$ and the mean square transmission coefficient $\langle |T|^2 \rangle = 1 - \langle |R|^2 \rangle$ can be found by employing (3.6). The expressions obtained agree exactly with those given in Refs. 5 and 6.

4. PERTURBATION ANALYSIS OF PROBLEM 2

In order to calculate moments of $u(x, L)$, the probability density function $p_2(\mathbf{Y}, x, t)$ will be needed. This p.d.f. satisfies the diffusion equation (2.22) with initial condition (2.24). The solution to this equation will now be found for small ϵ.

Before doing this, we simplify (2.22) by expanding

$p_2(\mathbf{Y},x,t)$ in terms of the eigenfunctions $\exp(in\theta + i\omega v)$, $n = 0, \pm 1, \pm 2, \cdots, -\infty < \omega < \infty$, i.e.,

$$p_2(\mathbf{Y},x,t) = \sum_{n=-\infty}^{+\infty} \int_{-\infty}^{+\infty} Q_n(\omega,\mathbf{X},x,t)e^{i(\omega v+n\theta)}d\omega. \quad (4.1)$$

Using (4.1) in (2.22) and the orthogonality properties of the eigenfunctions, we have

$$\frac{\partial Q_n}{\partial t} = \hat{L}_2 Q_n, \quad \hat{L}_2 = L_1^{(0)} + \epsilon\hat{L}_2^{(1)}, \quad x \le t \le L, \quad (4.2)$$

where $L_1^{(0)}$ is defined in (2.17) and where

$$\hat{L}_2^{(1)} = L_1^{(1)} + \tfrac{1}{2}k\mu[i\omega\rho \sin\phi - in(1 + \rho\cos\phi)] \quad (4.3)$$

with $L_1^{(1)}$ being given by (2.18). The initial condition for (4.2) can be obtained by inverting (4.1) with $t = x$, using (2.24), (2.25), (2.26) and simplifying. We have

$$Q_n(\omega,\mathbf{X},x,x) = (2\pi)^{-2}(1 + \rho e^{+i\phi})^{-(n+i\omega)/2}(1 + \rho e^{-i\phi})^{(n-i\omega)/2}p_1(\mathbf{X},x). \quad (4.4)$$

Since (4.2) is similar to the diffusion equation treated in Sec. 3, we again employ the perturbation procedure used there. By using this method we find an approximate expression for $Q_n(\omega,X,x,t)$ which is similar to (3.4). Putting this in (4.1), we obtain an approximate expression for the p.d.f. $p_2(X,x,t)$ over the interval $x \le t \le L$. Then integrating out the μ dependence, we finally obtain

$$\hat{p}_2(\mathbf{Y},x,t) = \int_{-\infty}^{+\infty} p_2(\mathbf{Y},x,t)d\mu \quad (4.5)$$

$$= \sum_{n=-\infty}^{+\infty} \int_{-\infty}^{+\infty} \hat{Q}_n(\omega,\hat{\mathbf{X}},x,t)e^{in\theta+i\omega v}d\omega, \quad (4.6)$$

where

$$\hat{Q}_n(\omega,\hat{\mathbf{X}},x,t) = \int_{-\infty}^{+\infty} Q_n(\omega,\mathbf{X},x,t)d\mu, \quad (4.7)$$

with $\hat{\mathbf{Y}} = (v,\theta,\rho,\phi)$ and $\hat{\mathbf{X}} = (\rho,\phi)$. The expression for $\hat{Q}_n$ is given by

$$\hat{Q}_n(\omega,\hat{X},x,t) = \sum_{n=-\infty}^{\infty} e^{im\phi}\int_0^{\infty} h_{mn}(s)M_{mn}(z)P^m_{-1/2+is}(z) \times e^{[-2imk+\gamma_{mn}\epsilon^2](t-x)}ds + O(\epsilon^2), \quad x \le t \le L, \quad (4.8)$$

where

$$M_{mn}(z) = (z-1)^{1/2+m}(z+1)^{(3+i\omega-4m-4)/2} \quad (4.9)$$

and

$$\gamma_{mn} = 8^{-1}(1 + 4k^2)^{-1}k^2[(2is + 2m + n)^2 - 1 - 2(2m + n)^2 + 4ik(2m + n) - 2(2m + n)^2(1 + 4k^2)]. \quad (4.10)$$

In the above z is given in terms of ρ by (3.7) and $P^{\mu}_{\nu}(z)$ is the associated Legendre function of degree ν and order μ. The coefficients $h_{mn}(s)$ appearing in (4.8) are found by setting $t = x$ and inverting this expression. We find

$$h_{mn}(s) = J_m(s)\int_{-\pi}^{+\pi}\int_1^{\infty}[M_{nm}(z')]^{-1} \times P^m_{-1/2+is}(z')e^{-im\phi'}\hat{Q}_n(\omega,\hat{\mathbf{X}},x,x)d\phi'dz', \quad (4.11)$$

where

$$J_m(s) = (2\pi^3)^{-1/2}s \sinh\pi s\Gamma(\tfrac{1}{2} - m + is)\Gamma(\tfrac{1}{2} - m - is). \quad (4.12)$$

The initial data $\hat{Q}(\omega,\hat{\mathbf{X}},x,x)$ is obtained by integrating both sides of (4.4) with respect to μ, then using (4.7) (3.5) and (3.6).

Although the expression derived above for $\hat{p}_2(\hat{\mathbf{X}},x,t)$ is fairly complex, we will see that expression for specific moments are much simpler.

5. MOMENTS OF $u(x, L)$

The mean $\langle u(x,L)\rangle$ and intensity $\langle |u(x,L)|^2\rangle$ will be calculated to $O(\epsilon^2)$ from our knowledge of $\hat{p}_2(\hat{\mathbf{X}},x,L)$. We first consider the mean of $u(x,L)$. Using (2.20), we have

$$\langle u\rangle = e^{ik(L-x)}\langle e^{v+i\theta}\rangle \quad (5.1)$$

$$= e^{ik(L-x)}\int e^{v+i\theta}\hat{p}_2(\hat{\mathbf{Y}},x,L)d\hat{\mathbf{Y}}. \quad (5.2)$$

Substituting (4.6) with $t = L$ into (5.2) and integrating with respect to v, θ, and ω gives

$$\langle u\rangle = (2\pi)^2e^{ik(L-x)}\int Q_{-1}(i,\hat{\mathbf{X}},x,L)d\hat{\mathbf{X}}. \quad (5.3)$$

We now use (4.8) and (5.3) and then we perform the ϕ integration. After this we change integration variables from ρ to z with the aid of (3.7). The result is

$$\langle u\rangle = (2\pi)^3e^{ik(L-x)}\int_1^{\infty}\int_0^{\infty} h_{0,-1}(s) \times P_{-1/2+is}(z)e^{\epsilon^2\delta_{0,-1}(s)(L-x)}dzds + O(\epsilon^2), \quad (5.4)$$

where

$$h_{0,-1}(s) = \frac{s\tanh\pi s}{(2\pi)^3}\int_1^{\infty}\int_0^{\infty} s'\tanh\pi s' \times P_{-1/2+is}(z')P_{-1/2+is'}(z') \times e^{\lambda_0(s')\epsilon^2L}dz'ds' \quad (5.5)$$

and

$$\delta_{0,-1}(s) = -8^{-1}(1 + 4k^2)^{-1}k^2[(2s + i)^2 + 3 + 4ik + 2(1 + 4k^2)]. \quad (5.6)$$

The double integrals appearing in (5.5) and (5.4) are evaluated in Appendix B using properties of the Mehler transform. We use (B5) to evaluate the double integral in (5.5). This result is then used in (5.4) and that double integral is evaluated with the aid of (B6). We find

$$\langle u\rangle = e^{ik(L-x)+\epsilon^2\delta_{0,-1}(-i/2)(L-x)+\epsilon^2\lambda_0(-i/2)\epsilon^2L} + O(\epsilon^2). \quad (5.7)$$

Now using (5.6), (3.7) and introducing the unnormalized variables in (2.3), we obtain

$$\langle u\rangle = e^{i\kappa(\bar{L}-\bar{x})} + O(\epsilon^2), \quad (5.8)$$

where

$$\kappa = ik - \tfrac{1}{2}\epsilon^2k^2l[(1 + 2k^2l^2 + ikl)/(1 + 4k^2l^2)]. \quad (5.9)$$

This result has been obtained by Kupiec, *et al.*[4] when they applied the method of smooth perturbation to the Dyson equation. It should be noted that our κ is the complex conjugate of their κ. This is because our waves are incident from opposite sides of the slab.

The intensity $\langle |u(x,L)|^2\rangle$ can be evaluated by using (2.20) and p.d.f. $\hat{p}_2$. We have

$$\langle |u|^2\rangle = \langle e^{2v}\rangle = \int e^{2v}\hat{p}_2(\hat{\mathbf{Y}},x,L)d\hat{\mathbf{Y}}. \quad (5.10)$$

Substituting (4.6) with $t = L$ into (5.10) and integrating with respect to v, θ, and ω gives

$$\langle |u|^2 \rangle = (2\pi)^2 \int \hat{Q}_0(2i, \hat{\mathbf{X}}, x, L) d\hat{\mathbf{X}}. \tag{5.11}$$

We use (4.8) in (5.11) and then we evaluate the ϕ integral. Following this, the ρ integration variable is transformed to z with the aid of (3.7). We obtain

$$\langle |u|^2 \rangle = (2\pi)^3 \int_1^\infty \int_0^\infty \frac{h_{0,0}(s)}{1+z} P_{-1/2+is}(z) e^{\lambda_0(s)\epsilon^2(L-x)} dz ds + O(\epsilon^2), \tag{5.12}$$

where

$$h_{0,0}(s) = \frac{2s \tanh \pi s}{(2\pi)^3} \int_1^\infty \int_0^\infty s' \tanh(\pi s') z' P_{-1/2+is}(z') \times P_{-1/2+is'}(z') e^{\lambda_0(s')\epsilon^2 L} dz' ds' \tag{5.13}$$

with $\lambda_0(s)$ given in (3.7). The recurrence relation[12]

$$(2\nu + 1) z P_\nu(z) = (\nu + 1) P_{\nu+1}(z) + \nu P_{\nu-1}(z) \tag{5.14}$$

for the Legendre function $P_\nu(z)$ can be used in (5.13). Two double integrals result which are given in (B7) and (B8) of Appendix B. We obtain

$$h_{0,0}(s) = [\tanh \pi s/(2\pi)^3][(2s - i) e^{\lambda_0(s-i)\epsilon^2 x} + (2s + i) e^{\lambda_0(s+i)\epsilon^2 x}]. \tag{5.15}$$

Now by putting (5.15) into (5.12) and evaluating the z integral with the aid of[13]

$$\int_1^\infty (z+1)^{-1} P_{-1/2+is}(z) dz = \pi \operatorname{sech} \pi s, \tag{5.16}$$

the desired expression for $\langle |u|^2 \rangle$ is obtained. This result is then expressed in terms of the unnormalized variables by using (2.3). We find

$$\langle |u|^2 \rangle = \pi e^{4\tilde{x} - \tilde{L}} \int_0^\infty \frac{\tanh \pi s}{\cosh \pi s} [\sin 8\tilde{x}s + 2s \cos 8\tilde{x}s] e^{-4s^2\tilde{L}} ds + O(\epsilon^2), \quad 0 \le \tilde{x} \le \tilde{L}, \tag{5.17}$$

where

$$\tilde{x} = d\epsilon^2 \bar{x}, \quad \tilde{L} = d\epsilon^2 \bar{L},$$
$$d = k^2 l / 8(1 + 4k^2 l^2). \tag{5.18}$$

Since we have not been able to evaluate this integral in terms of known functions, it has been evaluated numerically for various slab widths. These graphs are shown in Fig. 1. For small and large $\tilde{L}$ we find that (5.10) simplifies to

$$\langle |u|^2 \rangle = 1 + O(\epsilon^2), \quad 0 \le \tilde{x} \le \tilde{L}, \quad \tilde{L} \ll 1$$
$$\langle |u|^2 \rangle = \operatorname{erfc}[\tilde{L}^{1/2}(1 - 2\tilde{x}/L)] + O(\epsilon^2), \quad 0 \le \tilde{x} \le \tilde{L}/2, \quad \tilde{L} \gg 1, \tag{5.19}$$
$$= 2 - \operatorname{erfc}[\tilde{L}^{1/2}(1 - 2\tilde{x}/\tilde{L})] + O(\epsilon^2), \quad \tilde{L}/2 < \tilde{x} < \tilde{L} \tag{5.20}$$

where $\operatorname{erfc}(z)$ is the complementary error function of z. The approximate expression for large $\tilde{L}$ was obtained by using the saddle point method to evaluate (5.17). The complementary error function arose because of the presence of a pole near a saddle point.

As a check of the consistency of the result (5.17), we find $\langle |T|^2 \rangle$ by evaluating $\langle |u|^2 \rangle$ at $\tilde{x} = 0$. The expression for $\langle |T|^2 \rangle$ obtained agrees with that obtained in Ref. 6.

ACKNOWLEDGMENTS

The author would like to thank Professor J. B. Keller for bringing this problem to his attention and Professor R. Pickholtz for many helpful discussions. He also wishes to thank the National Science Foundation for supporting this work under Grant GK-27881.

APPENDIX A

To derive (2.9)-(2.11) from (2.8), we first take the derivative of (2.8) with respect to t and interchange orders of differentiation. We obtain

$$\left(\frac{\partial^2}{\partial x^2} + k^2[1 + \epsilon\mu(x)]\right) \frac{\partial u(x,t)}{\partial t} = 0, \tag{A1}$$

$$\left(\frac{\partial}{\partial x} + ik\right) \frac{\partial}{\partial t} u(0,t) = 0, \tag{A2}$$

$$\frac{\partial^2 u(t,t)}{\partial x^2} + \frac{\partial^2 u(t,t)}{\partial x \partial t} - ik \frac{\partial u(t,t)}{\partial x} - ik \frac{\partial u(t,t)}{\partial t} = 0. \tag{A3}$$

We denote $\partial u/\partial x$ and $\partial u/\partial t$ as the partial derivatives of u with respect to the first and second arguments of u respectively.

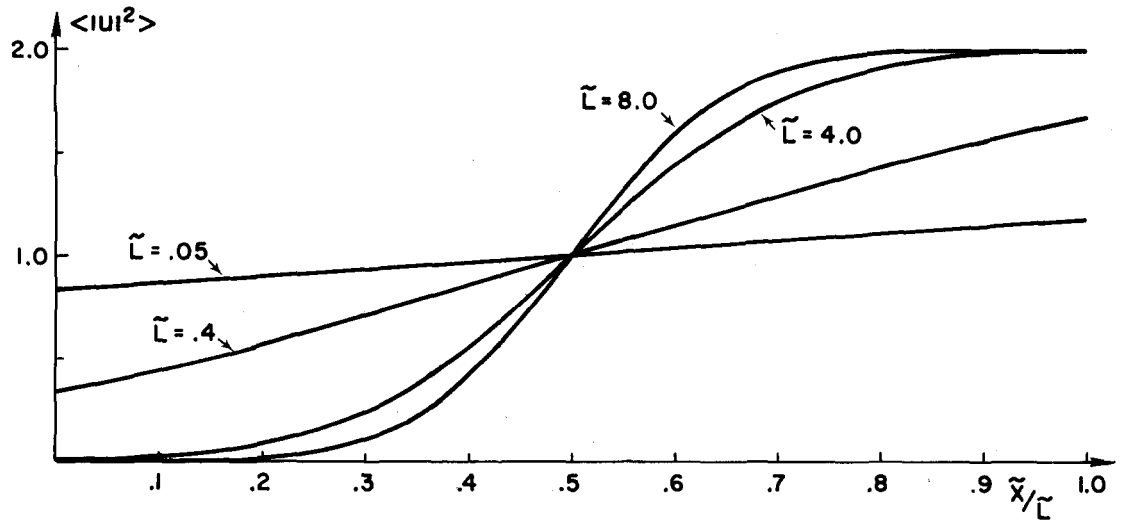


FIG. 1. Intensity vs $\tilde{x}/\tilde{L}$ with $\tilde{L}$ as a parameter.

An inspection of (A1), (A2), (2.8a), and (2.8b) shows that u and $\partial u/\partial t$ obey the same equation and the same boundary condition at $x = 0$. Thus we conclude that

$$\frac{\partial u(x,t)}{\partial t} = C(t)u(x,t). \tag{A4}$$

Now using (2.8a) and (A4) with $x = t$ in (A3), we obtain

$$C(t) = ik\{1 + \tfrac{1}{2}\epsilon\mu(t)[1 + \Gamma(t)]\} \tag{A5}$$

with

$$\Gamma(t) = \frac{iku(t,t) + (\partial/\partial x)u(t,t)}{iku(t,t) - (\partial/\partial x)u(t,t)}. \tag{A6}$$

We shall call $\Gamma(t)$ the generalized reflection coefficient since from (A6), (2.6c), and (2.7), we see that

$$\Gamma(L) = \frac{iku(L,L) + (\partial/\partial x)u(L,L)}{iku(L,L) - (\partial/\partial x)u(L,L)} = u(L) - 1 = R. \tag{A7}$$

An equation for $\Gamma(t)$ can be obtained by taking the derivative of (A6) with respect to t and then using (2.8a), (A4) with $x = t$ and (A6) in this expression. We find

$$\frac{d\Gamma(t)}{dt} = 2ik\Gamma(t) + \tfrac{1}{2}ik\epsilon\mu(t)[1 + \Gamma(t)]^2, \tag{A8}$$

where $\Gamma(0) = 0$. The above equation is a Ricatti equation. An equation of this type was found for the reflection coefficient in Ref. 5. The initial condition is found by making use of (2.8b) and (A6) with $x = t = 0$.

The reflection coefficient of the slab can thus be found by solving (A8) over the interval $0 \le t \le L$. The field $u(x, L)$ can be found by first solving (A8) in the interval $0 \le t \le x$ and then solving (A4) and (A8) simultaneously in the interval $x \le t \le L$. The initial conditions for the system of equations (A4) and (A8) are

$$\Gamma(t)\,|_{t=x} = \Gamma(x), \tag{A9}$$

$$u(x,x) = 1 + \Gamma(x). \tag{A10}$$

Here $\Gamma(x)$ is the solution to the Ricatti equation in the interval $0 \le t \le x$ with $t = x$. The initial condition (A10) is obtained by making use of (2.8c) and (A6) with $t = x$.

APPENDIX B

The purpose of this appendix is to evaluate several double integrals that appear in the text. We start by considering the identity

$$s\tanh\pi s\, e^{r(s)} = s\tanh\pi s \int_1^\infty P_{-1/2+is}(z)q(z)dz, \qquad 0 \le s \le \infty, \tag{B1}$$

where

$$q(z) = \int_0^\infty P_{-1/2+is'}(z)[s'\tanh\pi s' e^{r(s')}]ds'. \tag{B2}$$

Here $r(s)$ is a second degree polynomial in s with complex coefficients, i.e.,

$$r(s) = a_2 s^2 + a_1 s + a_0, \tag{B3}$$

where we assume $0 > \mathrm{Re}a_2 > \mathrm{Im}a_2$.

The identity (B1) follows immediately from the fact that we are taking the inverse Mehler transforms of the Mehler transform[14] of $s\tanh\pi s e^{r(s)}$.

Simplifying (B1), we have

$$e^{r(s)} = \int_1^\infty P_{-1/2+is}(z)q(z)dz. \tag{B4}$$

One can now show that both sides of equality (B4) can be analytically continued into an infinite strip in the complex s plane bounded by the lines $s = \pm 2i$. The analytic continuation is dependent on the fact that $q(z)$ must decay rapidly enough as $z \to \infty$ so that the integral in (B4) is uniformly convergent. We find $q(z)$ has this property when $0 > \mathrm{Re}a_2 > \mathrm{Im}a_2$.

We now use (B4) to evaluate the integrals of interest. The functions $\lambda_0(s)$ and $\delta_{0,-1}(s)$ used below are given in (3.7) and (5.6) respectively.

Case 1: Let $r(s) = \lambda_0(s)\epsilon^2 L$:

$$e^{\lambda_0(-i/2)\epsilon^2 L} = \int_1^\infty \int_0^\infty s'\tanh\pi s' P_{-1/2+is}(z') \times P_{-1/2+is'}(z')e^{\lambda_0(s')\epsilon^2 L}dz'ds'. \tag{B5}$$

Case 2: Let $r(s) = \delta_{0,-1}(s)\epsilon^2(L-x) + \lambda_0(s)\epsilon^2 L$ with $s = -i/2$:

$$e^{\delta_{-1}(-i/2)\epsilon^2(L-x)+\lambda_0(-i/2)\epsilon^2 L} = \int_1^\infty \int_0^\infty s\tanh\pi s P_{-1/2+is'}(z') \times e^{\delta_{0,-1}(s)\epsilon^2(L-x)+\lambda_0(s)\epsilon^2 L}dz'ds. \tag{B6}$$

Here $P_0(z) = 1$ was used.

Case 3: Let $r(s) = \lambda_0(s)\epsilon^2 L$ and replace s by $s - i$ in (B4):

$$e^{\lambda_0(s-i)\epsilon^2 L} = \int_1^\infty \int_0^\infty s'\tanh\pi s' P_{1/2+is}(z') \times P_{-1/2+is'}(z')e^{\lambda_0(s')\epsilon^2 L}dz'ds'. \tag{B7}$$

Case 4: Let $r(s) = \lambda_0(s)\epsilon^2 L$ and replace s by $s + i$ in (B4):

$$e^{\lambda_0(s+i)\epsilon^2 L} = \int_1^\infty \int_0^\infty s'\tanh\pi s' P_{-3/2+is}(z') \times P_{-1/2+is'}(z')e^{\lambda_0(s')\epsilon^2 L}dz'ds'. \tag{B8}$$

[1] L. Breiman, *Probability* (Addison-Wesley, New York, 1968), p. 347.
[2] M. E. Gertsenshtein and V. B. Vasil'iev, Theor. Prob. Appl. 4, 392 (1959).
[3] Yu. L. Gazaryan, Zh. Eksp. Teor. Fiz. 56, 1856 (1969) [Sov. Phys. JETP 29, 996 (1969)].
[4] I. Kupiec, L. B. Felsen, S. Rosenbaum, J. B. Keller, and P. Chow, Radio Sci. 4, 1067 (1969).
[5] G. C. Papanicolaou, SIAM J. Appl. Math. 21, 13 (1971).
[6] J. A. Morrison, G. C. Papanicolaou, and J. B. Keller, Comm. Pure Appl. Math. 24, 473 (1971).
[7] R. Bellman and R. Kalaba, J. Math. Mech. 8, 683 (1959).
[8] R. Bellman and R. Kalaba, *Electromagnetic Wave Propagation* (Academic, New York, 1960), p. 243.
[9] E. Wong, *Stochastic Processes in Information and Dynamical Systems* (McGraw-Hill, New York, 1972).
[10] A. Jazwinski, *Stochastic Processes and Filtering Theory* (Academic, New York, 1970).
[11] Since the transformation (2.14) is singular at $\rho = 0$, any initial distribution of ϕ can be chosen.
[12] W. Magnus, F. Oberhettinger, and R. P. Soni, *Formulas and Theorems for the Special Functions of Mathematical Physics* (Springer-Verlag, Berlin, 1966), p. 171.
[13] Bateman Manuscript Project, *Higher Transcedental Functions* (McGraw-Hill, 1953), Vol. 1, p. 174.
[14] See Ref. 10, p. 398.

Some rigorous results for the vertex model in statistical mechanics

H. J. Brascamp* and H. Kunz†

Institut des Hautes Etudes Scientifiques, 91 Bures-sur-Yvette, France

F. Y. Wu‡

Department of Physics, Northeastern University, Boston, Massachusetts 02115

(Received 28 June 1973)

It is shown that the free and periodic boundary conditions are completely equivalent for the ice-rule (six-vertex) models in zero field. With an external direct or staggered field, we establish that in an ice-rule model the free and periodic boundary conditions are equivalent, and also equal to some special boundary conditions, either at sufficiently low temperatures or with sufficiently high fields in the appropriate direction. Regions of constant direct polarization are found. We also establish the existence of the spontaneous staggered polarization in an antiferroelectric using the Peierls argument.

I. INTRODUCTION

We consider in this paper some outstanding unsolved problems in the vertex models in statistical mechanics.[1] One problem whose solution has proven to be elusive in the past is the proof of the equivalence of the free and periodic boundary conditions for the six-vertex (ice-rule) models. In the solution of the ice-rule models obtained by Lieb,[2] it is crucial that a periodic boundary condition is used. Attempts in trying to show that the free boundary condition would yield the same solution have not succeeded.[3] Another problem which has not been considered before is the proof of the existence of a long-range order in an antiferroelectric. Such a proof is useful and desirable, especially since the only available calculation is that of the F model and is based on assumptions that appear to be difficult to justify.[4] In this paper we consider both of these problems.

In Sec. II we define the various vertex models and standardize the notation to be used in the ensuing proofs. The equivalence of various boundary conditions for the six-vertex models is considered in Sec. III using the weak-graph expansions. In particular, it is established that with no external field the free and periodic boundary conditions are completely equivalent. With an external direct and staggered field, we establish that the free and periodic boundary conditions are equivalent, and also equal to some special boundary conditions consistent with the energetically favored configurations, either at sufficiently low temperatures or with sufficiently high fields. These results, while expected, have not previously been proved. In Sec. IV, using the Peierls argument, we establish the existence of the spontaneous staggered polarization in a general antiferroelectric, including the F model. Several challenging unsolved problems related to the F model are presented in Sec. V.

II. BASIC DEFINITIONS

Let L be a two-dimensional square lattice of $N = n \times n$ vertices, with n even. The lattice edges of L which terminate in two vertices of L are called interior edges and those terminating in one vertex called exterior. A vertex which is the terminus of one or more exterior edges is called a boundary vertex and the set of such vertices is called the boundary of L. The vertex model is defined by placing arrows on the edges of L.

We follow the notation of Ref. 1 in defining the vertex models. The most general vertex model that can be defined on L is the 16-vertex model. The 16 arrow configurations that can occur at a vertex are shown in Fig. 1.

Let e_ξ be the energy associated with the vertex of type $\xi (= 1, 2, \ldots, 16)$ and $\xi(i)$ be the type of configuration of the ith vertex. The partition function of the 16-vertex model is

$$Z = \sum \prod_{i=1}^{N} \omega_{\xi(i)} \tag{1}$$

where $\omega_\xi = \exp(-e_\xi/kT)$ is the Boltzman factor and the summation is extended to all arrow (or bond) configurations subject to a given boundary condition.

If all 16 ω's are nonzero, we have a 16-vertex problem. Otherwise we have an eight-vertex model if ω_ξ is nonzero for $\xi = 1, 2, \ldots, 8$ only, and a six-vertex problem if $\omega_\xi \neq 0$ for $\xi = 1, 2, \ldots, 6$. Some physical properties of the six-vertex models appear to be rather different from the corresponding eight- or 16-vertex ones, and it is the main purpose of this paper to stress and discuss them.

We study a general six-vertex model defined by[1]

$$\begin{aligned} &e_1 = \epsilon_1 - (h + v), \quad e_2 = \epsilon_1 + h + v, \\ &e_3 = \epsilon_2 - h + v, \quad e_4 = \epsilon_2 + h - v, \\ &e_5 = \epsilon_3 + s, \quad e_6 = \epsilon_3 - s \quad \text{for vertices on sublattice } A, \end{aligned} \tag{2}$$

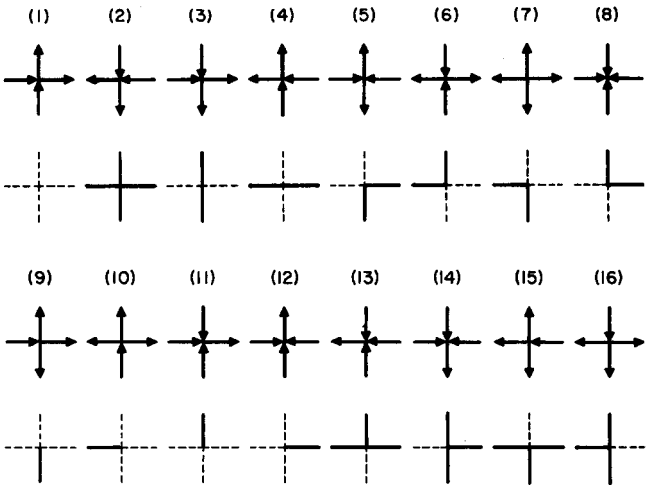


FIG. 1. The vertex configurations of the 16-vertex model and the associated bond configurations using the convention C_2.

$e_5 = \epsilon_3 - s, \quad e_6 = \epsilon_3 + s$ for vertices on sublattice B,

other $e_\xi = \infty$.

The electric fields included are the horizontal and vertical direct fields h and v, and the staggered quadrupole field s.

The general KDF model is defined by

$$\epsilon_2 > \epsilon_1, \quad \epsilon_3 > \epsilon_1. \tag{3}$$

In the absence of any field, the ground state configurations are X_1 and X_2, which correspond to complete filling of the lattice with vertices $\xi = 1, 2$, respectively. An equivalent definition is $\epsilon_1 > \epsilon_2$, $\epsilon_3 > \epsilon_2$, with ground states X_3 and X_4.

The general F model is defined by

$$\epsilon_1 > \epsilon_3, \quad \epsilon_2 > \epsilon_3. \tag{4}$$

For sufficiently small direct fields and without staggered field, the ground state configurations are X_{56} and X_{65}, where X_{56} is given by

$\xi(i) = 5$, i on sublattice B,

$\xi(i) = 6$, i on sublattice A.

X_{65} is obtained from X_{56} by interchanging 5 and 6. The same ground states are found for the generalization of the F model to the 16-vertex antiferroelectric

$$e_5 = e_6 = 0, \quad e_\xi > 0 \quad \text{for all } \xi \neq 5, 6. \tag{5}$$

We shall study the dependence of the free energy

$$\mathfrak{F} = -kT \lim_{N\to\infty} (1/N) \ln Z \tag{6}$$

on the boundary conditions. A boundary condition (BC) on the finite lattice L is expressed by a restriction on the arrow directions on the exterior edges of L. We speak of free boundary conditions (FBC) if there is no restriction, of periodic boundary conditions (PBC) if the two sets of horizontal (resp. vertical) exterior edges have identical arrow configurations. If the directions of the arrows on the exterior edges are all specified, we have a special boundary condition (SBC). We shall use SBC S_α, $\alpha = 1, 2, 3, 4, 56, 65$; to denote the boundary condition fitting to the configuration X_α. In Figs. 2a, b, S_1 and S_{56} are pictured. The boundary condition under which $\mathfrak{F}$ is defined will be denoted by the subscript F, P, or α.

As a consequence of the ice rule, the SBC S_α, $\alpha = 1, \ldots, 4$, completely determines the configuration X_α in the interior of L. Therefore, if the free energy (6) satisfies

$$\mathfrak{F}_F = \mathfrak{F}_\alpha$$

for $\alpha = 1, \ldots, 4$, then the system is in the frozen state X_α; in particular, the direct polarization is saturated in the corresponding direction.

Furthermore, the SBC S_{56} and S_{65}, which allow equal numbers of in and out arrows along the exterior rows and columns, only permit configurations in L with total direction polarization zero. Indeed, by the ice rule the vertical polarization is conserved from row to row, the horizontal from column to column. Therefore, if

$$\mathfrak{F}_F = \mathfrak{F}_{56} \text{ or } \mathfrak{F}_{65},$$

in some region in (h, v), the direct polarization is zero in that region.

Since the free energy is different for the different frozen states, we already see that the free energy cannot be independent of the BC. This difficulty disappears as soon as the weights satisfy

$$\omega_1, \ldots, \omega_8 > 0, \quad \omega_9, \ldots, \omega_{16} \geq 0. \tag{7}$$

In that case, a volume $n \times n$ with any BC can be imbedded in a volume $(n+2) \times (n+2)$ with any other BC, the energy difference being of the order of n. This implies then that the limiting free energy (6) does not depend on the BC (see, e.g., Ref. 5, Lemma 2.2.1).

It is useful to introduce representations of the vertex configurations in terms of bond graphs. The bond graphs in Fig. 1 are obtained by drawing bonds for each $\leftarrow$ and $\downarrow$ arrow. Under this convention, which we shall call C_2, vertex 2 becomes the one with four bonds, configuration X_2 becomes X_B with bonds everywhere, and SBC S_2 becomes SBC S_B with bonds on all exterior edges. Analogously, we have the convention C_α, $\alpha = 1, \ldots, 4$, 56, 65, which carries X_α over into X_B and S_α into S_B.[6] In the bond language we also meet SBC S_H, Fig. 1c, in which no bonds (holes) appear on all exterior edges. In Sec. III the bond representations are used, together with the weak graph transformation, to prove the equivalence of boundary conditions in various cases.

In Sec. IV we use a representation of the arrow configurations in terms of closed polygons. For the 16-vertex antiferroelectric (5), we prove the existence of the spontaneous staggered polarization

$$P_0 = -\left(\frac{\partial \mathfrak{F}}{\partial s}\right)_{s=0+} \tag{8}$$

for sufficiently low temperatures. For the F model we also find equivalence of FBC to SBC S_{56} or S_{65}, which gives regions, where the direct polarization is zero and the staggered polarization is independent of the direct field.

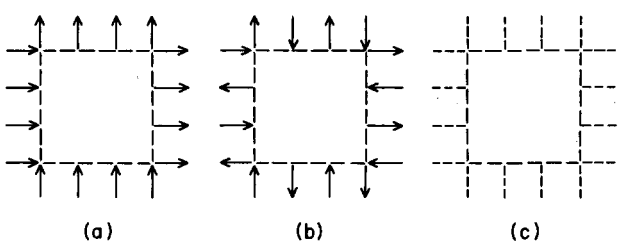


FIG. 2. Special boundary conditions for a 4×4 lattice. (a) SBC S_1, (b) SBC S_{56}, (c) SBC S_H. The upper left vertex belongs to sublattice B.

III. EQUIVALENCE OF BOUNDARY CONDITIONS FOR THE ICE-RULE MODELS

Our proof is based on the application of the weak-graph expansion.[7] It was first shown by Nagle[8] that under the weak-graph expansion the six-vertex KDP and F models are transformed into eight-vertex models. For a general lattice model the weak-graph expansion will lead to a 16-vertex model. If all the transformed vertex weights are positive, one can then establish the equivalence of the boundary conditions as indicated above. The crux of matter is therefore to find a weak-graph expansion which will generate positive weights.

It will be convenient for our discussions to first review briefly the formulation of the weak-graph expansion. For simplicity we shall use bond graphs. For each edge connecting vertices i and j we introduce a matching factor

$$\tfrac{1}{2}[1 + c_{ij}(\xi_i)c_{ij}(\xi_j)] = 1, \quad \text{if } \xi_i \text{ and } \xi_j \text{ are compatible,}$$
$$= 0, \quad \text{otherwise,} \tag{9}$$

where

$$c_{ij}(\xi_i) = c_{ji}(\xi_i) = 1, \quad \text{if } \xi_i \text{ has a bond on edge } ij,$$
$$= -1, \quad \text{otherwise.} \tag{10}$$

Then the partition function takes the form

$$Z = \sum_{\xi_i=1}^{16} \prod_{\substack{\text{matching} \\ \text{edges}}} \tfrac{1}{2}[1 + c_{ij}(\xi_i)c_{ij}(\xi_j)] \prod_{i=1}^{N} \omega(\xi_i). \tag{11}$$

For FBC there are $2n(n-1)$ matching (interior) edges and for PBC there are $2n^2$, and for SBC S_B, $2n(n+1)$. Note that the matching factors for the exterior edges in the case of S_B have the form $\frac{1}{2}[1 + c_{ij}(\xi_i)]$.

Next we expand the product of the edge factors. Each term in the expansion is now a product of many $c_{ij}(\xi_i)c_{ij}(\xi_j)$ factors which can be conveniently represented graphically by drawing bonds between the connected vertices i and j. After rearrangement, (11) can be rewritten as

$$Z = c \sum_{G} \prod_{i=1}^{N} \left(\tfrac{1}{4} \sum_{\xi_i=1}^{16} \omega(\xi_i) \prod_{\substack{k \\ ik \text{ in } G}} c_{ik}(\xi_i) \right), \tag{12}$$

where the summation is extended to all bond graphs G, the constant $c = 1$ for PBC, $c = 2^{2n}$ for FBC and $c = 2^{-2n}$ for SBC S_B.

The summation inside the square bracket in (12) can be considered to be some new vertex weights ω'_ξ defined by the bond graphs G. Since the bonds in G are drawn only on the matched edges, which include the exterior edges for S_B and do not include them for FBC, we find

$$Z_P(\omega) = Z_P(\omega'), \quad Z_F(\omega) = 2^{2n} Z_H(\omega'),$$
$$Z_B(\omega) = 2^{-2n} Z_F(\omega'). \tag{13}$$

The new weights are

$$\omega'(\xi_i) = \tfrac{1}{4} \sum_{\eta_i} \omega(\eta_i) \prod_{\substack{k \\ ik \text{ in } \xi_i}} c_{ik}(\eta_i). \tag{14}$$

Detailed expression of this linear transformation can be found in Eq. (408) of Ref. 1 and will not be reproduced here.

If the transformed weights ω'_ξ satisfy (7), then the thermodynamic limit (6) is independent of the BC. For the original weights ω_ξ we have then, by (13),

$$\mathfrak{F}_P(\omega) = \mathfrak{F}_F(\omega) = \mathfrak{F}_B(\omega). \tag{15}$$

Let us apply this basic idea to various six-vertex models.

(i) General six-vertex model in zero field: Here we show the equivalence $\mathfrak{F}_P = \mathfrak{F}_F$ for the general six-vertex model in zero field. Let $u_i = \exp(-\epsilon_i/kT)$, $i = 1, 2, 3$. Further, we take the fields $h = v = s = 0$ in (2). With convention C_2, the weak-graph expansion leads to the eight-vertex problem

$$\omega'_1 = \omega'_2 = \tfrac{1}{2}(u_1 + u_2 + u_3), \quad \omega'_3 = \omega'_4 = \tfrac{1}{2}(u_1 + u_2 - u_3),$$
$$\omega'_5 = \omega'_6 = \tfrac{1}{2}(u_1 - u_2 + u_3), \quad \omega'_7 = \omega'_8 = \tfrac{1}{2}(u_1 - u_2 - u_3),$$
$$\text{other } \omega'_\xi = 0. \tag{16}$$

Some of these weights may be negative, but if n is even the partition functions for PBC and SBC S_H are invariant under the reversal of sign of these weights.[9] Then we conclude from (13) that

$$\mathfrak{F}_P(\omega) = \mathfrak{F}_F(\omega)$$

at all temperatures. Strictly speaking, the above reasoning does not hold for the special temperature, where two of the transformed weights are zero; but there the result follows by continuity of the free energies in the temperature. The equality of $\mathfrak{F}_P$ and $\mathfrak{F}_F$ to $\mathfrak{F}_B$ does not follow generally, because the FBC partition function of the transformed weights is not invariant under the reversal of sign. For $T < T_c$ we also establish in the following that $\mathfrak{F}_P(\omega) = \mathfrak{F}_F(\omega) = \mathfrak{F}_\alpha(\omega)$, where X_α is the ground state configuration.

(ii) General six-vertex model in direct and staggered fields: For the general six-vertex model in nonzero fields, we now show that $\mathfrak{F}_F = \mathfrak{F}_P = \mathfrak{F}_\alpha$, $\alpha = 1, 2, 3, 4, 56, 65$, provided that the direct or the staggered field is large enough in the appropriate direction.

Consider the general six-vertex model (2). We put $H = h/kT$, $V = v/kT$, $t = s/kT$. With convention C_2, the weak-graph expansion leads to the weights

$$\omega'_1, \ldots, \omega'_8 = \tfrac{1}{2}[u_1 \cosh(H + V) \pm u_2 \cosh(H - V) \pm u_3 \cosh t],$$
$$\omega'_9, \ldots, \omega'_{16} = \tfrac{1}{2}[-u_1 \sinh(H + V) \pm u_2 \sinh(H - V) \pm u_3 \sinh t], \tag{17}$$

each combination of the signs occurring twice. The transformed weights ω' satisfy the positivity condition (7) if

$$u_1 \cosh(H + V) > u_2 \cosh(H - V) + u_3 \cosh t,$$
$$-u_1 \sinh(H + V) \geq u_2 \sinh|H - V| + u_3 \sinh|t|. \tag{18}$$

This may always be satisfied at a given temperature by taking $h \approx v \ll -|s|$, independent of the other parameters. So then we have

$$\mathfrak{F}_F(\omega) = \mathfrak{F}_P(\omega) = \mathfrak{F}_2(\omega). \tag{19}$$

By using the convention S_α, $\alpha = 1, \ldots, 4$, we find in the same way that

$$\mathfrak{F}_F(\omega) = \mathfrak{F}_P(\omega) = \mathfrak{F}_\alpha(\omega) \tag{20}$$

if the direct field is large enough in the appropriate direction. The system is then in the frozen state X_α with saturated direct polarization. Similarly if the vertex (2) is favored, i.e., $e_2 < e_\alpha$, $\alpha \neq 2$, then (18), and hence (19), are always satisfied at sufficiently low temperatures. The same conclusion holds for (20) from which we conclude that if the vertex energies are such that the vertex (α), $\alpha = 1, \ldots, 4$, is favored, the system is in the frozen state X_α at sufficiently low temperatures, a result borne out by the exact solution.[1]

In the case that the configuration 56 is favored, we use the convention C_{56} so that the weak-graph transformation gives the weights

$$\omega_1', \ldots, \omega_8' = \tfrac{1}{2}[u_3 \cosh t \pm u_1 \cosh(H + V) \pm u_2 \cosh(H - V)],$$
$$\omega_9', \ldots, \omega_{16}' = \tfrac{1}{2}[u_3 \sinh t \pm u_1 \sinh(H + V) \pm u_2 \sinh(H - V)]. \quad (21)$$

Using (7), this leads to the conditions

$$u_3 \cosh t > u_1 \cosh(H + V) + u_2 \cosh(H - V),$$
$$u_3 \sinh t \geq u_1 \sinh|H + V| + u_2 \sinh|H - V|. \quad (22)$$

By taking the staggered field $s \gg |h| \approx |v|$, at a given temperature, this can always be satisfied. Then

$$\mathfrak{F}_F(\omega) = \mathfrak{F}_P(\omega) = \mathfrak{F}_{56}(\omega). \quad (23)$$

With convention C_{65}, we find that

$$\mathfrak{F}_F(\omega) = \mathfrak{F}_P(\omega) = \mathfrak{F}_{65}(\omega) \quad (24)$$

if $s \ll -|h| \approx -|v|$. In both cases, the direct polarization is zero. Similarly if the vertex energies are such that the configuration 56 is favored and $s > 0$, then (22), hence (23), is always satisfied at sufficiently low temperatures. Likewise, we find (24) to hold at sufficiently low temperatures and $s < 0$ if the configuration 65 is favored. The proof breaks down for $s = 0$ and h or $v \neq 0$ (the F model in a direct field). Fortunately an alternate proof, which is valid for the F model with *arbitrary* s, exists. For continuity or reading, details of the latter proof will be given in the Appendix. We remark that, under (23), (24), or (A1), the free energy is independent of the direct field h and v. It follows that, in particular from (A1), P_0 *is independent of* h *and* v.

(iii) The KDP and F models: While the results in (ii) above are sufficiently general, it is illuminating to specialize these conclusions to some special cases:

KDP model in zero field: The exact transition temperature T_c is known to be given by $u_1 = u_2 + u_3$; hence the condition (18) corresponds to $T < T_c$. We conclude that the system is in a frozen state for $T < T_c$, a result known from the exact solution.[1] More generally we have

$$\mathfrak{F}_F(\omega) = \mathfrak{F}_P(\omega) = \mathfrak{F}_\alpha(\omega), \; T < T_c \quad (25)$$

where X_α, $\alpha = 1, 2, 3, 4$, is the ground state configuration.

KDP model in a vertical field:—The condition (18) for the system to be in a frozen state X_2 becomes

$$u_1 \cosh V > u_2 \cosh V + u_3, \quad v \leq 0. \quad (26)$$

It can be easily checked that the temperature given by (26) is lower than the exact transition temperature T_c [cf. (327) in Ref. 1]. Consequently the conclusion derived here is somewhat weaker than the known exact result[1] which states that the system is frozen for $T < T_c$.

F model in a zero field: we conclude that

$$\mathfrak{F}_F(\omega) = \mathfrak{F}_P(\omega) = \mathfrak{F}_{56}(\omega) = \mathfrak{F}_{65}(\omega), \quad T < T_c, \quad (27)$$

where the transition temperature T_c is given by the relation $u_3 = u_1 + u_2$.[1]

F model in a staggered field: For $s > 0$ we find from (22) that (23) holds if

$$u_3 \cosh t > u_1 + u_2. \quad (28)$$

We conjecture that (28) yields a lower bound on the exact transition temperature for a nonzero staggered field. If also a direct field v is present, Eqs. (18), (22), and (A2) determine regions of constant polarization 0, ± 1 in the (v, T) plane, which have the same qualitative shape as the regions found for $s = 0$ (Ref. 1, Fig. 31).

IV. EXISTENCE OF P_0

In this Section we use Peierls argument[10] to establish the existence of the spontaneous staggered polarization P_0 in the general antiferroelectric (5).[11] We need only to show the existence for the SBC S_{56}. For the eight- and 16-vertex models the result is independent of the BC. For the F model, we then use the equivalence of SBC S_{56} with other BC established in Sec. III in conjunction with the standard concavity argument[12] to extend the existence of P_0 to FBC and PBC.

The first step of our proof is to introduce for a vertex model a graph representation consisting of closed contours. Consider a lattice L' composed of all lattice points of L and also the intersecting points of the diagonals in L. The edges of L' are the half-diagonals in L. An example of L' and its relationship with L is shown in Fig. 3 for a 4×4 lattice. Consider the four quadrants belonging to a vertex of L. If a quadrant is bounded by two arrows pointing in or by two arrows pointing out, we draw a bond along the half-diagonal bisecting the quadrant. Then one can easily see that there are always an even number of bonds meeting at any interior vertex of L'. This is also the case on the boundary of L', as illustrated in Fig. 3, if one assumes SBC S_{56} or S_{65} in L. The result is then a *one-to-one correspondence between the arrow configurations on* L (assuming S_{56} or S_{65}) *and the closed-polygonal configuration on* L'.[13] This correspondence holds even if some of the vertex energies of L are $+\infty$ such as in the F model. This only restricts the allowed contours on L' and will not affect the following discussions.

With these preliminaries, it is now relatively easy to formulate the Peierls argument. Since our proof follows closely the standard argument,[12] we shall only point out the essential points. From (8) we have

$$P_0 = \lim_{n \to \infty} (1/N)\langle n_+ - n_- \rangle,$$
$$n_+ = n_{5B} + n_{6A}, \quad n_- = n_{5A} + n_{6B}, \quad (29)$$

where n_{5B} denotes the number of (5) vertices on sublattice B, etc., and $\langle\ \rangle$ is the thermal average. Denote the vertices of n_+ by "+" and n_- by "—", as shown in Fig. 3, then the contours on L' separate seas of "+" from "seas" of "—". In fact, the contours go through all vertices except (5) and (6). Furthermore, the boundary of L' cannot be "—". Let

$$\epsilon = \min\{e_1, \ldots, e_4, \tfrac{1}{2}e_7, \tfrac{1}{2}e_8, e_9, \ldots, e_{16}\} > 0. \quad (30)$$

Then each bond segment on L' has a weight of at most $\exp(-\epsilon/2kT)$. The standard argument then gives

$$(1/N)\langle n_- \rangle \leq \sum_{b \geq 4} (b/4)^2 \, 3^{b-1} e^{-b\epsilon/2kT}. \quad (31)$$

Similarly the average fraction of vertices other than (5) and (6) is bounded by

$$1 - (1/N)\langle n_+ + n_- \rangle \leq \sum_{b \geq 4} 3^{b-1} e^{-b\epsilon/2kT}. \quad (32)$$

The right-hand sides of (31) and (32) can be made arbitrarily small at sufficiently low temperatures. It follows then $(1/N)\langle n_+ - n_- \rangle$ is nonzero and hence $P_0 > 0$ exists. This completes the proof.

V. DISCUSSIONS

Our methods are especially suited for discussion of the low temperature properties of the vertex models. They do not allow us to establish the equivalence of FBC and PBC for the six-vertex models at high temperature, except in the case of zero field where the existence of extra symmetries permits us to complete the proof.

We wish to point out that it is a mere consequence of the ice rule that the six-vertex model can be put in a frozen state X_α, $\alpha = 1, 2, 3, 4$, if the vertex α is favored. This frozen state disappears as soon as the ice rule is violated. Indeed, another way to state these properties is to say that the free energy is not a strictly convex function of the direct fields h and v at sufficiently low temperatures. However, it has been shown[14] that the eight- or the 16-vertex model in a direct field h, v and a staggered field is equivalent to an Ising model on a square lattice with finite, but short-ranged, interactions. Recently, Griffiths and Ruelle[15] have shown that the free energy of such lattice system is a strict convex function of the translationally invariant interactions. Since the translationally invariant interactions of the corresponding Ising model depend linearly on h and v, the free energy of the eight- or 16-vertex models will be a strictly convex function of h and v. This then rules out the occurrence of the frozen states in these models.

To conclude our discussion, we list some unsolved problems related to the F model which appear to be particularly challenging:

(i) Are $\mathfrak{F}_F$ and $\mathfrak{F}_P$ analytic in s near $s = 0$ at sufficiently high temperatures? We note that Baxter[16] has shown $\mathfrak{F}_P$ is singular at $s = 0$ when $T = 2T_c$.

(ii) What is the decay of the correlation functions? Is it exponential like in an ordinary lattice gas with short-ranged interactions?

(iii) We expect $P_0 = 0$ for $T > T_c$ and a rigorous proof of this fact appears to be lacking.

(iv) In the F model, like in an antiferromagnet, the translational invariance is broken at low temperatures and under small direct fields. In the (v, T) plane, for example, we then expect a boundary B_1 which separates the regions where $P_0 = 0$ and $P_0 \neq 0$, as in an ordinary antiferromagnet. Does this boundary coincide with the boundary B_2 separating the region of vertical polarization $y = 0$ and $y \neq 0$? Note that B_1, but not B_2, disappears if there is a nonzero staggered field. On the other hand, as discussed in the above, it is B_2, not B_1, which disappears when one goes over to an eight-vertex case.

(v) Does the staggered susceptibility diverge along the boundary B_1 separating regions of different long-range orders. Here we note that Baxter found that the staggered susceptibility diverges at $T = 2T_c$!

ACKNOWLEDGMENT

We would like to thank Dr. N. H. Kuiper for the hospitality extended to us at the Institut des Hautes Etudes Scientifiques where this research was performed.

APPENDIX

In this appendix we prove the following result for the F model specified by (2) and (4):

$$\mathfrak{F}_F(s) = \mathfrak{F}_P(s) = \mathfrak{F}_{56}(s), \quad s \geq 0,$$
$$= \mathfrak{F}_{65}(s), \quad s \leq 0, \tag{A1}$$

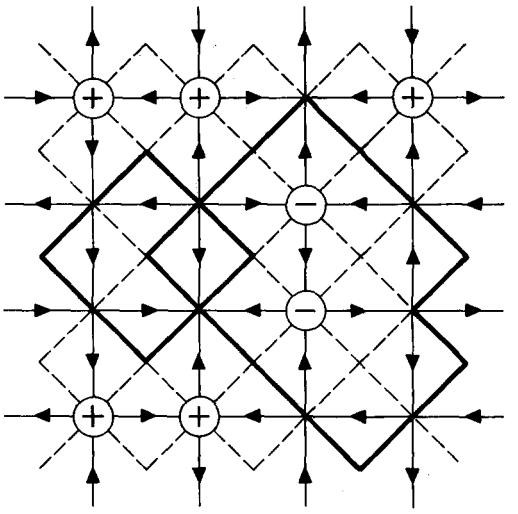


FIG. 3. Relationship between the lattices L and L'. Arrow configurations on L are mapped into closed-polygonal configurations on L'. The mapping is one-to-one if L assumes the SBC S_{56} as shown here. The polygons on L' separates the + and − vertices on L and the upper left vertex belongs to sublattice B.

provided that

$$T \leq \min\{T_c, (\epsilon + |s|)/\ln 3\}. \tag{A2}$$

Here T_c is the transition temperature of the F model in zero field, and ϵ is the minimum vertex energy defined by (29). First we establish the following lemma:

Lemma:

$$\left.\begin{aligned} Z_{56} &\geq Z_{65}, \quad s \geq 0 \\ Z_{65} &\geq Z_{56}, \quad s \leq 0 \end{aligned}\right\}, T \leq T_c.$$

Proof: The partition functions Z_{56} and Z_{65} are independent of h and v. Consider Z_{56} and Z_{65} as functions of the complex variable $z = \exp(s/kT)$. The boundary condition S_{56} can be simulated by taking $z = \infty$ at the boundary of the volume. It then follows from the work of Suzuki and Fisher[17] (see also Chang *et al.*[18]), that the partition function belongs to the Lee–Yang class (Ref. 17, Def. 1) for $T \leq T_c$, so that then we have[19]

$$Z_{56}(z) \neq 0, \quad |z| \geq 1. \tag{A3}$$

Also by arrow reversal we have, for all z,

$$Z_{56}(z) = Z_{65}(z^{-1}), \tag{A4}$$

which says that, for $|z| = 1$,

$$Z_{56}(z) = Z_{65}(z^*) = Z_{65}{}^*(z). \tag{A5}$$

Hence the function

$$f(z) \equiv Z_{65}(z)/Z_{56}(z)$$

is analytic in z for $|z| \geq 1$ ($z = \infty$ included) and satisfies

$$|f(z)| = 1 \quad \text{for } |z| = 1. \tag{A6}$$

The maximum modulus principle then gives

$$|f(z)| \leq 1 \quad \text{for all } |z| \geq 1. \tag{A7}$$

In particular this leads to the first part of the Lemma. The second part is obtained by symmetry.

Now we proceed to prove (A1). First consider $s \geq 0$. Using the notation of Sec. IV, we note that with FBC on L the contours can terminate at the boundary of L'. So any configuration can be described in terms of connected paths λ_k beginning and ending at the boundary together with closed contours. The paths λ_k divide L into subvolumes Λ_j which have the BC S_{56} or S_{65} (adjacent subvolumes have opposite BC). Since the polygon-to-arrow-configurations correspondence is two-to-one, then to each path $\{\lambda_k\}$ on L' there correspond two terms of the form

$$\prod_k e^{-u(\lambda_k)/kT} \prod_j Z_\alpha(\Lambda_j) \tag{A8}$$

in the partition function $Z_F(L)$. Here α stands for either 56 or 65 and $u(\lambda_k)$ is the sum of the vertex energies along the path λ_k. Thus, by the lemma just proved, an upper bound on $Z_F(L)$ results for $s \geq 0$ if we replace all Z_α in (A8) by Z_{56}. Furthermore, we note that

$$Z_{56}(L) \geq \prod_k e^{s|\lambda_k|/kT} \prod_j Z_{56}(\Lambda_j) \tag{A9}$$

since every term on the right-hand side of (A9) appears in $Z_{56}(L)$ but not vice versa; here $|\lambda_k|$ is the number of vertex points along λ_k. Using (29), we have also

$$u(\lambda_k) \geq \epsilon|\lambda_k|. \tag{A10}$$

Combining (A8)–(A10) and replacing Z_α by Z_{56} in (A8), we then obtain the bound

$$Z_F \leq Z_{56} D, \quad s \geq 0, \tag{A11}$$

where

$$D = \sum_{\{\lambda_k\}} e^{-(\epsilon+s)|\lambda_k|/kT}. \tag{A12}$$

Now each λ_k must pass through at least one point on the boundary, we obtain

$$D \leq \sum_{\alpha=0}^{4n+4} \binom{4n+4}{\alpha} \left(\sum_{b \geq 1} 3^b e^{-b(\epsilon+s)/kT} \right)^\alpha = \left(1 + \sum_{b \geq 1} 3^b e^{-b(\epsilon+s)/kT} \right)^{4n+4}, \tag{A13}$$

where $4n + 4$ is the number of vertices on the boundary of L'. The series in (A13) converges if

$$kT < (\epsilon + s)/\ln 3. \tag{A14}$$

On the other hand, we have generally[1]

$$Z_{56} \leq Z_P \leq Z_F. \tag{A15}$$

Thus we deduce the first statement ($s \geq 0$) of (A1) on taking the thermodynamic limits of (A11) and (A15). The statement with $s \leq 0$ can be similarly obtained or by symmetry considerations.

We remark that the result (A1) is valid for arbitrary s. Furthermore, for small $|s|$, the temperature bound (A2) is better than the one obtained in Sec. III using the weak-graph expansions.

* On leave of absence from the University of Groningen, The Netherlands. Supported by The Netherlands Organization from the Advancement of Pure Research (Z.W.O.). Present address: Institute for Advanced Study, Princeton, N.J.
† Supported by Swiss National Foundation for Scientific Research. Present address: Laboratoire de Physique Theorique, Lausanne, Switzerland.
‡ Supported in part by the National Science Foundation.

[1] E. H. Lieb and F. Y. Wu, in *Phase Transitions and Critical Phenomena*, Vol. I, edited by C. Domb and M. S. Green (Academic, London, 1972).
[2] E. H. Lieb, Phys. Rev. Letters **18**, 692, 1046 (1967); **19**, 108 (1967).
[3] See discussions in Ref. 1.
[4] R. J. Baxter, J. Phys., C, L94 (1973) and private communications.
[5] D. Ruelle, *Statistical Mechanics* (Benjamin, New York, 1969).
[6] The mapping C_{56} is that given in Fig. 5 of Ref. 1.
[7] J. F. Nagle, J. Math. Phys. **9**, 1007 (1968).
[8] J. F. Nagle, J. Math. Phys. **7**, 1492 (1966).
[9] C. Fan and F. Y. Wu, Phys. Rev. B **2**, 723 (1970).
[10] R. Peierls, Proc. Cambr. Phil. Soc. **32**, 477 (1936); R. B. Griffiths, Phys. Rev. 136, A437 (1964); R. L. Dobrushin, Teor. Veroyat. Primen. **10**, 209 (1965) [Theory Prob. Appl. **10**, 193 (1965)]; Funkt. Anal. Pril. 2, 44 (1968) [Funct. Anal. Appl. **2**, 302 (1968)].
[11] If the vertex energies satisfy certain symmetry such as in an eight-vertex model in zero field, this result then also establishes the existence of a spontaneous polarization of a ferroelectric in a direct field in 45° direction in the first quadrant. See discussions in H. J. Brascamp, H. Kunz, and F. Y. Wu, J. Phys. C. L164 (1973).
[12] See, for example, discussion by R. B. Griffiths, in Ref. 1.
[13] The correspondence is two-to-one for periodic lattices.
[14] Pp. 348-53, Ref. 1.
[15] R. B. Griffiths and D. Ruelle, Comm. Math. Phys. **23**, 169 (1971).
[16] R. J. Baxter, Phys. Rev. B **1**, 2199 (1970).
[17] M. Suzuki and M. E. Fisher, J. Math. Phys. **12**, 235 (1970).
[18] K. S. Chang, S. Y. Wang, and F. Y. Wu, Phys. Rev. A **4**, 2324 (1971).
[19] The conditions of Theorem 1 in Ref. 17 imply that $Z(z)$ here is of the *Lee–Yang class*.

An exact solution of a scalar-tensor theory of gravitation*

E. Ihrig and D. K. Sen

Department of Mathematics, University of Toronto, Toronto M5S, 1A1, Canada
(Received 16 May 1973)

An exact, closed-form, static spherically symmetric solution to the field equations of a scalar-tensor theory is given. This is compared with the series solution given previously. Also, an analysis is made of the geodesics of the theory.

1. INTRODUCTION

Recently a new scalar-tensor theory[1,2] of gravitation was proposed by one of the present authors. A formal series solution of the static spherically symmetric case was also given at the same time. In the meantime Halford[3] has given a closed-form solution in *isotropic* coordinates, and also analyzed the *autoparallels* of the affine connection. He concluded thereby that the theory predicts the same effects, within observational limits, as the Einstein theory. We pointed out,[1] however, that in this theory the geodesics of the metric are not identical with the autoparallels of affine connection—they differ in the numerical values of some of the coefficients involved in the two equations. Since it is more natural to assume that test particles follow geodesics rather than autoparallels, it seems worthwhile to analyze the geodesics also. In this paper we give an exact closed-form solution in the *standard* Schwarzschild coordinates and compare with the previously given series solution. We also analyze some of the geodesics.

2. AN EXACT STATIC SPHERICALLY SYMMETRIC SOLUTION

We shall follow the notations of Sen and Dunn.[1] The field equations (in vacuum) are

$$R_{\alpha\beta} - \tfrac{1}{2} g_{\alpha\beta} R + \omega (x^0)^{-2} x^0_{,\alpha} x^0_{,\beta} + (\omega/2)(x^0)^{-2} g_{\alpha\beta} x^0_{,\nu} x^{0,\nu} = 0 \tag{2.1}$$

where $\omega = 3/2$.

In the static spherically symmetric case, we have

$$ds^2 = (x^0)^2 (e^{\nu} dt^2 - e^{\lambda} dr^2 - r^2 d\theta^2 - r^2 \sin^2\theta \, d\phi^2), \tag{2.2}$$

where

$$\lambda = \lambda(r), \quad \nu = \nu(r), \quad x^0 = x^0(r).$$

Equation (2.1) thus reduces to the following equations

$$e^{\nu} = D + C\phi(r), \tag{2.3}$$

$$e^{\lambda} = Ar^4 [\phi'(r)]^2 / [D + C\phi(r)], \tag{2.4}$$

$$x^0 = \text{const} \exp \int \{ - [(4/\omega r^2) + (2/\omega r)(\phi''/\phi')] \}^{1/2} dr, \tag{2.5}$$

where ϕ satisfies the following differential equation:

$$(D + C\phi)(1 + r\phi''/\phi') + Ar^4 \phi'^2 - Cr\phi' = 0. \tag{2.6}$$

Here A, C, D are integration constants. Our main purpose is to solve (2.6). If we make the substitution $v = r\phi'/(D + C\phi)$ in (2.6) we have $v'/v^2 = -Ar\phi' = 2Ar\phi - (r^2\phi)'$ and thus

$$(C/2)(-v^{-1} + Ar^2\phi)' = CAr\phi \tag{2.7}$$

and

$$(-v^{-2}/2 + \tfrac{1}{2} ADr^2)' = -CAr\phi. \tag{2.8}$$

Adding (2.7) to (2.8) we get the first integral

$$1 + Cv - (ADr^2 + ACr^2\phi - K)v^2 = 0, \tag{2.9}$$

where K is an integration constant. Consider first two special cases.

$C = 0$:

From (2.9) we have

$$\phi' = \pm D/r(ADr^2 + K)^{1/2}. \tag{2.10}$$

Substituting in (2.3), (2.4), and (2.5) we get

$$e^{\nu} = D, \quad e^{\lambda} = ADr^2/(ADr^2 + K), \tag{2.11}$$

$$x^0 = \text{const} \exp \int \left(- \frac{2K}{\omega r^2 (ADr^2 + K)} \right)^{1/2} dr. \tag{2.12}$$

Integrating (2.10) we find that $\phi(r) = -(1/\sqrt{-K}) \sin^{-1}[\sqrt{-K/A}(1/r)]$ if $D = 1$. This agrees with the series solution given by Sen and Dunn[1] for $r > |K/A|^{1/2}$. As far as we know, this solution has no analogue in the Brans–Dicke[4] theory.

$K = 0$:

From (2.9) we have $v = 1/(Ar^3\phi' - C)$, which together with (2.7) gives $\phi' = K_2/r^2$, $\phi = -K_2/r + K_3$, where $AK_2^3 = (D + CK_3)K_2$ in view of (2.6). Thus $x^0 =$ const and we obtain the Schwarzschild case.

$C \neq 0, K \neq 0$:

From (2.7) we have $rv'/v = -Ar^3\phi' v$ which on substitution in (2.9) gives $-dr/r = dv/v(1 + Cv - Kv^2)$ or

$$r = L \, [\exp(- \textstyle\int dv/v(1 + Cv - Kv^2)], \tag{2.13}$$

where L is an integration constant. And

$$\phi = [M(\exp C \textstyle\int [- dv/(1 + Cv - Kv^2]) - D]/C. \tag{2.14}$$

Equations (2.13) and (2.14) give implicitly the exact solution of (2.6). They can be simplified in certain cases. For example, if $K < 0$, $C^2 + 4K > 0$, we have

$$\phi = (M\lambda^{2n+1} - D)/C, \tag{2.15}$$

where $sL(a_- - a_+)\lambda^n - ra_+\lambda + ra_- = 0$,

$$n = \tfrac{1}{2}(C/\sqrt{C^2 + 4K} - 1), \quad a_{\pm} = (-C \pm \sqrt{C^2 - 4K})/2K,$$

$s = \text{sgn}[(a_+ - a_-)/(1 - \lambda)]$. In evaluating (2.15) from (2.13) and (2.14) we had put $\lambda = (v + a_+)/(v + a_-)$.

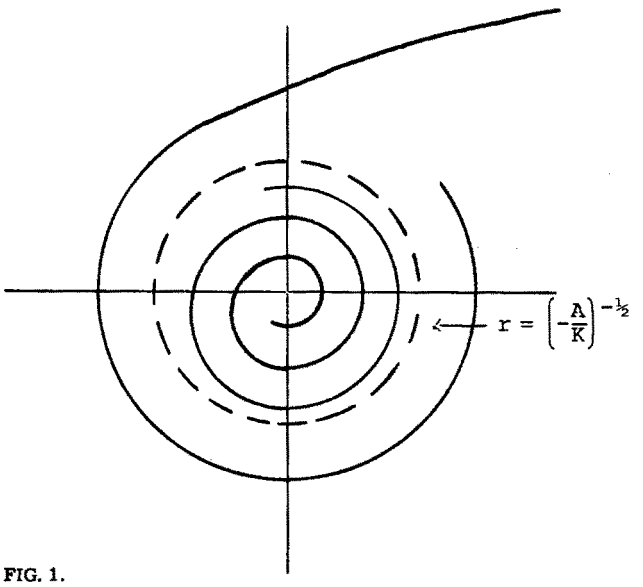


FIG. 1.

From (2.13) and (2.14) one can show that there exists a number r_0 such that the series expansion $\phi(r) = \sum_{n=0}^{\infty} a_n r^{-n}$ is valid for $r > r_0$ and which agrees with the one previously given.

Note that in this theory $ds^2 = (x^0)^2 g_{\mu\lambda} dx^\mu dx^\lambda = \bar{g}_{\mu\lambda} dx^\mu dx^\lambda$, where $\bar{g}_{\mu\lambda} = (x^0)^2 g_{\mu\lambda}$. It is possible to calculate the first few terms of g_{00}, for example. We have

$$\bar{g}_{00} = 1 - (|C|/\sqrt{A} + \sqrt{8/\omega}\,\sqrt{-K/A})(1/r) + 0(1/r^2). \tag{2.16}$$

3. GEODESICS

The geodesics are given by

$$\delta \int \left((x^0)^2 g_{\mu\kappa} \frac{dx^\mu}{d\sigma} \frac{dx^\kappa}{d\sigma}\right)^{1/2} d\sigma = 0. \tag{3.1}$$

For time-like geodesics σ would be the proper time s. Equation (3.1) has as a first integral ($dx^\mu/d\sigma = \dot{x}^\mu$)

$$K_1 = (x^0)^2[e^{\nu(r)}\dot{t}^2 - e^{\lambda(r)}\dot{r}^2 - r^2\dot{\theta}^2 - r^2 \sin^2\theta\dot{\phi}]. \tag{3.2}$$

For motion in the equatorial plane $\dot{\theta} = 0$, $\theta = \pi/2$. The relevant Euler–Lagrange equations are then

$$[(x^0)^2 e^{\nu(r)}\dot{t}]^{\cdot} = 0, \tag{3.3}$$

$$[(x^0)^2 r^2 \sin^2\theta\dot{\phi}]^{\cdot} = 0 \tag{3.4}$$

or

$$\dot{t} = C_1/(x^0)^2 e^{\nu(r)}, \quad \dot{\phi} = C_2/(x^0)^2 r^2. \tag{3.5}$$

where C_1, C_2 are constants.

Substituting in (3.2) we get

$$\dot{r}^2 = e^{-\lambda(r)}[C_1^2/(x^0)^4 e^{2\nu(r)} - C_2^2/(x^0)^4 r^2 - K_1/(x^0)^2] \tag{3.6}$$

or

$$\sigma - \sigma_0 = \int e^{\lambda(r)/2} x^0 dr/[C_1^2/(x^0)^2 e^{2\nu(r)} - C_2^2/(x^0)^2 r^2 - K_1]^{1/2}. \tag{3.7}$$

Also from (3.5)

$$\phi - \phi_0 = \int C_2 \frac{e^{\lambda(r)/2}}{r}[C_1^2 r^2 - e^{-2\nu(r)} - C_2^2 - K_1(x^0)^2 r^2]^{1/2} dr. \tag{3.8}$$

Consider now the special case $C = 0$, $D = 1$. Then

$$\dot{t} = C_1/(x^0)^2, \quad \dot{\phi} = C_2/(x^0)^2 r^2, \quad \theta = \pi/2 \tag{3.9}$$

and

$$\phi - \phi_0 = \int \frac{C_2\sqrt{A}\, dr}{[-K_1(x^0)^2 r^2 + C_1^2 r^2 - C_2^2]^{1/2}(Ar^2 + K)^{1/2}}. \tag{3.10}$$

The null geodesics are given by $K_1 = 0$. If we assume that $C_1^2/-C_2^2 = A/K$, we get

$$\phi - \phi_0 = \int \frac{Q' dr}{r^2 + K/A}$$

and

$$r = \begin{cases} Q' \tanh(Q(\phi - \phi_0)) & \text{if } r < (-A/K)^{-1/2} \\ Q' \coth(Q(\phi - \phi_0)) & \text{if } r > (-A/K)^{-1/2}, \end{cases} \tag{3.11}$$

where Q, Q' are constants. Fig. 1 shows the trajectory of light rays in the $\theta = \pi/2$ plane.

*Supported in part by the National Research Council of Canada.

[1]D. K. Sen and K. A. Dunn, J. Math. Phys. **12**, 578 (1971).
[2]D. K. Sen and J. R. Vanstone, J. Math. Phys. **13**, 990 (1972).
[3]W. D. Halford, J. Math. Phys. **13**, 1699 (1972).
[4]C. Brans and R. H. Dicke, Phys. Rev. **124**, 925 (1961).

Quantization of a general dynamical system

S.I. Ben-Abraham and A. Lonke

Department of Physics, University of the Negev, 84120 Beer-Sheva, Israel
(Received 5 April 1973)

Invariance implies that the momentum operator should be represented by the covariant derivative. For a general dynamical system, both covariant canonical and path integral quantization yield in the Schrödinger equation a correction term proportional to the curvature. Earlier erroneous calculations are hereby corrected and reconciled.

I. INTRODUCTION

De Witt[1] and, more recently, Cheng[2] have shown that the Schrödinger equation for a general dynamical system contains an additional term. This term is proportional to the scalar curvature, R, of the Riemannian configuration space whose metric is defined by the kinetic energy. When the curvature is not constant this term changes the energy spectrum.

The curvature term has been quite a puzzle. "Usual" quantization through commutation relations and "Feynman" quantization through path integrals[3] have so far yielded different and rather ambiguous results.

In this paper we endeavor to reconsider the question. We will show that a careful analysis leads to the same answer in both aforementioned cases.

The tensorial concepts and notations will lean heavily on Schouten[4] and Veblen,[5] which should be consulted.

II. COVARIANT COMMUTATION RELATIONS

We will show that, in a curved space, p_α, the canonically conjugate momentum operator to the coordinate q^α must be taken as $-i\hbar\nabla_\alpha$, where ∇_α is the operator of covariant derivation.

On the basis of classical mechanics, we postulate validity of the canonical commutation relations and deduce the correct form of p_α.

The commutation relations are

$$[q^\alpha, q^\beta] = 0, \tag{1a}$$

$$[p_\alpha, p_\beta] = 0, \tag{1b}$$

$$[q^\alpha, p_\beta] = I^\alpha_\beta, \tag{1c}$$

where q^α, p_β are canonical coordinates and momenta, I^α_β is the unit tensor, and brackets denote the commutator.

It is understood that:

(A) The q's and the p's are linear operators on a Hilbert space of functions ψ and so are their commutators.

The wave function ψ is defined on the $(n+1)$-dimensional state space $\{q^\alpha, t \mid \alpha = 1, \ldots, n\}$ (where t is time) which is assumed to be Riemannian. Then $\psi = \psi(q^\alpha, t)$ must be a density of weight 1/2 in the Riemannian configuration space $\{q^\alpha \mid \alpha = 1, \ldots, n\}$. Indeed, the volume element

$$dq = \underset{\sim}{\epsilon}_{\alpha_1 \ldots \alpha_n} dq^{\alpha_1} \ldots dq^{\alpha_n} \tag{2}$$

is manifestly a capacity (density of weight -1); here

$$\underset{\sim}{\epsilon}_{\alpha_1 \ldots \alpha_n} \triangleq \underset{\sim}{\epsilon}_{[\alpha_1 \ldots \alpha_n]}, \qquad \underset{\sim}{\epsilon}_{1, \ldots, n} \triangleq +1, \tag{3}$$

is the Levi-Civita capacity. But $\psi^* \psi dq$ must be a scalar. Hence ψ must be a density of weight 1/2; we shall call it a "half-density" for short.

The commutation relations (1) together with the postulate (A) imply that

$$p_\alpha = -i\hbar D_\alpha, \tag{4}$$

where the operator D_α stands for some kind of derivative with respect to q^α. In Euclidian space and Cartesian coordinates (c) it is simply the partial derivative,

$$D_c = \partial_c \triangleq \frac{\partial}{\partial q^c}. \tag{5}$$

However, given any vector v^α in a more general space, $\partial_\alpha v^\beta$ won't be a tensor. Yet, p_α is manifestly a covector in $\{q^\alpha\}$, whence $p_\alpha v^\beta$ must be a tensor. Consequently, we must postulate that:

(B) The operator p_α has the properties of a covector when applied to a tensor.

This leaves us no choice but to identify D_α with the covariant derivative ∇_α. Thus, we find

$$p_\alpha = -i\hbar\nabla_\alpha. \tag{6}$$

It must be remarked that the coordinates q^α are not the components of a vector. Neither are they genuine scalars. But the covariant derivative of the geometric object q^α is a tensor, namely

$$\nabla_\alpha q^\beta = \partial_\alpha q^\beta = I^\beta_\alpha. \tag{7}$$

This is the reason that it is unusual to write $\nabla_\alpha q^\beta$. It must be kept in mind, however, that the covariant derivative has been originally introduced to be consistent with (7).

Also, it is important to remember that the explicit form of the operator ∇_α depends on the kind of geometric object it is applied to.

III. LAGRANGIAN FORMALISM

We define the Lagrangian

$$L \triangleq \tfrac{1}{2} I_{\alpha\beta} \dot{q}^\alpha \dot{q}^\beta + A_\alpha \dot{q}^\alpha - V, \tag{8}$$

where the generalized mass tensor $I_{\alpha\beta}$, the vector potential A_α, and the scalar potential V are functions of the points (q^α, t) of state space, while the velocities are functions of time, t.

Introducing

$$ds^2 = I_{\alpha\beta} dq^\alpha dq^\beta \tag{9}$$

as the fundamental metric form in configuration space $\{q^\alpha\}$ the latter becomes a Riemannian space with $I_{\alpha\beta}$ as

metric tensor and $I^{\alpha\beta}$ as its inverse, which will be assumed to exist. The notation is motivated by the identity

$$I_{\alpha\kappa}I^{\kappa\beta} = I^{\beta}_{\alpha}, \tag{10}$$

whence it is seen that the metric tensor and its inverse are, in essence, the unit tensor with both indices lowered or raised, respectively.

The momentum is by definition

$$p_\alpha \triangleq \partial L/\partial \dot{q}^\alpha = I_{\alpha\beta}\dot{q}^\beta + A_\alpha. \tag{11}$$

Hence, the metric allows us to pass from the $\dot{q}$'s to the p's. In fact, $\dot{q}^\alpha$ and $p_\alpha - A_\alpha$ become the vector and covector components of the same object.

From (8) the Hamiltonian becomes, as usual,

$$\begin{aligned} H &= \tfrac{1}{2}(I^{\alpha\kappa}p_\kappa - A^\alpha)I_{\alpha\beta}(I^{\beta\lambda}p_\lambda - A^\beta) + V \\ &= \tfrac{1}{2}p_\alpha I^{\alpha\beta}p_\beta - A^\alpha p_\alpha + \tfrac{1}{2}A^\alpha A_\alpha + V, \end{aligned} \tag{12}$$

where

$$A^\alpha \triangleq I^{\alpha\beta}A_\beta. \tag{13}$$

The potential terms in L and H are irrelevant for our purpose and will be henceforth omitted.

IV. CURVATURE TERM BY CANONICAL QUANTIZATION

Now, we take the Hamiltonian and replace the momenta p_α by the operators (6)

$$H = \tfrac{1}{2}p_\alpha I^{\alpha\beta}p_\beta = -\frac{\hbar^2}{2}\nabla_\alpha I^{\alpha\beta}\nabla_\beta = -\frac{\hbar^2}{2}I^{\alpha\beta}\nabla_\alpha\nabla_\beta. \tag{14}$$

The last equality is valid because $I^{\alpha\beta}$ is covariant constant:

$$\nabla_\alpha I^{\beta\gamma} = 0. \tag{15}$$

In order to write out explicitly the expected extra term in the Hamiltonian we introduce new differential operators called extensions.[4,5] At the point $q = \overset{0}{q}$ of configuration space we introduce normal coordinates (i) with $\overset{0}{q}$ as origin $(q^i = 0)$ and t as a parameter. In normal coordinates, the affine connection Γ^m_{kl} vanishes at the origin while its partial derivatives behave as tensors:

$$\Gamma^m_{kl}(\overset{0}{q}) = 0, \tag{16}$$

$$\partial_{i_1\dots i_r}\Gamma^m_{kl} = N_{i_1\dots i_r kl}{}^{m}. \tag{17}$$

Now, $N_{i_1\dots kl}{}^{m}$ can be transformed back to general coordinates (α) by

$$N_{\alpha_1\dots\alpha_r\beta\gamma}{}^{\delta} = I^{i_1\dots i_r kl\delta}_{\alpha_1\dots\alpha_r\beta\gamma m}N_{i_1\dots i_r kl}{}^{m}. \tag{18}$$

$N_{i_1\dots i_r kl}{}^{m}$ is called the rth normal tensor; it is intimately connected with the curvature tensors up to the rth.

Any quantity can be "extended" by taking its partial derivative in normal coordinates (i) and afterwards transforming back to general coordinates (α) by a tensor transformation. For instance, if v^α is a vector, its rth extension is

$$\overset{*}{\nabla}_{\alpha_1\dots\alpha_r}v^\beta \triangleq I^{i_1\dots i_r\beta}_{\alpha_1\dots\alpha_r\kappa}\partial_{i_1\dots i_r}v^k. \tag{19}$$

The extensions differ from the covariant derivatives by terms containing only the quantity itself and powers of the normal tensors. The first extension is, of course, identical with the covariant derivative, but this is not true for the higher extensions.

After this digression, we are ready to evaluate H as given by (14). We apply the second covariant derivative $\nabla_{\alpha\beta}$ to the wave function ψ

$$\begin{aligned} \nabla_{\alpha\beta}\psi &= \nabla_\alpha(\partial_\beta\psi - \tfrac{1}{2}\Gamma_\beta\psi) \\ &= \partial_{\alpha\beta}\psi - \tfrac{1}{2}\psi\partial_\alpha\Gamma_\beta - \tfrac{1}{2}\Gamma_\beta\partial_\alpha\psi \\ &\quad - \Gamma^\kappa_{\alpha\beta}\partial_\kappa\psi - \tfrac{1}{2}\Gamma_\alpha\partial_\beta\psi + \tfrac{1}{2}\Gamma^\kappa_{\alpha\beta}\Gamma_\kappa\psi + \tfrac{1}{4}\Gamma_\alpha\Gamma_\beta\psi, \end{aligned} \tag{20}$$

where

$$\Gamma_\alpha \triangleq \Gamma^\kappa_{\alpha\kappa} = \tfrac{1}{2}\partial_\alpha \ln g, \tag{21}$$

and

$$g \triangleq \det(I_{\alpha\beta}) \tag{22}$$

is a density of weight 2.

In normal coordinates the Γ^m_{kl} vanish, Hence,

$$\begin{aligned} \nabla_{\alpha\beta}\psi &= \sqrt{\Delta}\,I^{ij}_{\alpha\beta}\{\partial_{ij}\psi - \tfrac{1}{2}\psi\partial_i\Gamma_j\} \\ &= \sqrt{\Delta}\,I^{ij}_{\alpha\beta}\{\partial_{ij}\psi - \tfrac{1}{2}N_{ij}\psi\} \\ &= \overset{*}{\nabla}_{\alpha\beta}\psi - \tfrac{1}{2}N_{\alpha\beta}\psi, \end{aligned} \tag{23}$$

where

$$\Delta \triangleq \det(I^\alpha_i), \tag{24}$$

$$N_{\alpha\beta} \triangleq N_{\alpha\beta\kappa}{}^{\kappa}. \tag{25}$$

We have

$$N_{\alpha\beta\gamma}{}^{\delta} = \tfrac{2}{3}R_{\alpha(\beta\gamma)}{}^{\delta}, \tag{26}$$

where $R_{\alpha\beta\gamma}{}^{\delta}$ is the (first) curvature tensor. In a Riemannian space

$$N_{\alpha\beta} = -\tfrac{1}{3}R_{\alpha\beta}, \tag{27}$$

is the Ricci tensor.

Thus, we derive

$$\nabla_{\alpha\beta}\psi = \overset{*}{\nabla}_{\alpha\beta}\psi + \tfrac{1}{6}R_{\alpha\beta}\psi. \tag{29}$$

Finally, for the Hamiltonian (14) we obtain

$$H\psi = -\frac{\hbar^2}{2}I^{\alpha\beta}\nabla_{\alpha\beta}\psi = -\frac{\hbar^2}{2}I^{\alpha\beta}\overset{*}{\nabla}_{\alpha\beta}\psi - \frac{\hbar^2}{12}R\psi, \tag{30}$$

where

$$R \triangleq I^{\alpha\beta}R_{\alpha\beta} \tag{31}$$

is the scalar curvature.

V. CURVATURE TERM BY PATH INTEGRATION

We consider the evolution of a dynamical system from time $t = t_0 - \tau$ up to time t_0. We wish to apply Feynman's method of path integration to a curved space. In doing so extreme care must be taken to keep all integrands scalars. Otherwise the integrals are not valid. Since the wave function ψ is a half-density one has to multiply it by factor $g^{-1/4}$ in order to obtain a scalar.

We introduce the abbreviations

$$f \triangleq f(q(t), t), \qquad f_0 \triangleq f(q(t_0), t_0). \tag{32}$$

The correct expression representing the evolution of the system becomes now

$$\underset{0}{\psi} = A\underset{0}{g}^{1/4} \int_{\{q\}} \exp\{iS/\hbar\}(\psi g^{-1/4})(g^{1/2}dq), \tag{33}$$

where S is the classical action

$$S \triangleq \underset{0}{S}(q,q) \triangleq \min \int_t^{t_0} L(\dot{q}(s), q(s))ds, \tag{34}$$

with boundary conditions

$$q(s=t) = q, \qquad \underset{0}{q}(s=t_0) = \underset{0}{q}. \tag{35}$$

A is a normalization factor to be determined later. Since L is a constant along a geodesic line, we have

$$S = L\tau = \underset{0}{L}\tau = \tfrac{1}{2}\underset{0}{I}_{\alpha\beta}\underset{0}{\dot{q}}^{\alpha}\underset{0}{\dot{q}}^{\beta}\tau. \tag{36}$$

In the limit $\tau \to 0$ (33) yields the Schrödinger equation. The integrals appearing in (33) are generalizations of the Gauss integral: They take the form[6]

$$\begin{aligned} &J^{\alpha_1\cdots\alpha_{2m}} \\ &\triangleq \int_{-\infty}^{\infty} dq^1 \cdots \int_{-\infty}^{\infty} dq^n \exp\left(\frac{i}{2\hbar\tau} I_{\alpha\beta}q^{\alpha}q^{\beta}\right)(q^{\alpha_1}\cdots q^{\alpha_{2m}}) \\ &= (2\pi i\hbar\tau)^{n/2}(i\hbar\tau)^m g^{-1/2}(2m-1)!! \\ &\quad \times I^{(\alpha_1\alpha_2}I^{\alpha_3\alpha_4}\cdots I^{\alpha_{2m-1}\alpha_{2m})}. \end{aligned} \tag{37}$$

Integrals with an odd number of q^{α}'s vanish, of course. Clearly, only a hypersphere of radius $(\hbar\tau/g)^{1/2}$ around the origin contributes to these integrals. Thus, one can expand the integrand around $\underset{0}{q}$ as a power series in $q^{\alpha} - \underset{0}{q}^{\alpha}$.

The expansions are best performed in normal coordinates (i) with origin $\underset{0}{q}{}^i = 0$. Then we have $q^i(t) = \dot{q}^i t$ and[4]

$$\begin{aligned} S &= \frac{\tau}{2}\{\underset{0}{I}_{ji} + \underset{0}{N}_{klji}q^k q^l + \underset{0}{N}_{mklji}q^m q^k q^l \\ &\quad + \underset{0}{N}_{nmklji}q^n q^m q^k q^l + \cdots\}\dot{q}^j\dot{q}^i \\ &= \frac{1}{2\tau}\{\underset{0}{I}_{ji} + \underset{0}{N}_{kl ji}q^k q^l + \cdots\}q^j q^i \\ &= \frac{1}{2\tau}\{\underset{0}{I}_{ji} + \frac{1}{3}\underset{0}{R}_{ljki}q^l q^k + \cdots\}q^j q^i. \end{aligned} \tag{38}$$

We will expand only up to the second term. However, we will show qualitatively how the higher terms enter into the Schrödinger equation.

For the wave function in normal coordinates we have the expansion

$$\begin{aligned} \psi &= \underset{0}{\psi}(q,t) + \frac{1}{1!}q^i\underset{0}{\nabla}_i\psi + \frac{1}{2!}q^j q^i\underset{0}{\nabla}_{ji}\psi + \cdots \\ &= \underset{0}{\psi}(q,t) + q^i\underset{0}{\partial}_i\psi + \tfrac{1}{2}q^j q^i(\underset{0}{\partial}_{ji} - \tfrac{1}{2}\underset{0}{N}_{ji})\psi + \cdots, \end{aligned} \tag{39}$$

while $g^{1/4}$ is constant in these coordinates.

We now substitute the expansions (38), (39) into (33). We leave the first term of (38) in the exponential and expand the rest. In Eq. (33) we expand ψ_0, evaluate the integrals on the rhs and determine the normalization factor A. This procedure yields

$$\begin{aligned} \psi + \tau\left(\frac{\partial\psi}{\partial t}\right) + \cdots &= \psi + \tfrac{1}{2}i\hbar\tau\Gamma^{ji}(\partial_{ji} - \tfrac{1}{2}N_{ji})\psi \\ &\quad - \tfrac{3}{2}i\hbar\tau I^{(ik}\Gamma^{ji)}N_{lkji}\psi + \cdots, \end{aligned} \tag{40}$$

where we have already omitted the subscript 0.

Now we apply (26), (27), and (31) to replace the normal tensors by curvature, transform Eq. (40) back to general coordinates, and, after multiplication by $i\hbar/\tau$ we finally obtain the Schrödinger equation,

$$i\hbar\frac{\partial\psi}{\partial t} = -\frac{\hbar^2}{2}I^{\alpha\beta}\overset{*}{\nabla}_{\alpha\beta}\psi - \frac{\hbar^2}{12}R\psi.$$

The equation is manifestly invariant. Clearly, the Hamiltonian is identical to that given by (30) which was obtained by canonical quantization.

VI. CONCLUSIONS AND OUTLOOK

For a generalized dynamical system, i.e., one whose kinetic energy form induces a nonvanishing and variable Riemannian curvature in configuration space, the Schrödinger equation contains an additional term proportional to the scalar curvature R. This curvature may be due to curvature of space-time or constraints. We have shown that both canonical and path integral quantization yield the same result, namely $+\hbar^2R/12$. This was achieved by making sure that the path integral (33) was invariant. This point was overlooked in previous calculations and was the source of errors and inconsistencies. It should be emphasized that the correction is of a kinematic nature and belongs to the kinetic energy. Therefore, it does not show up in the Lagrangian with reversed sign.

In most cases the curvature term will be insignificant. It might be interesting for highly condensed matter such as neutron stars. Of course, in this case the validity of the Schrödinger equation itself is questionable. Yet, if R is induced by constraints the term might well be significant and change the whole spectrum of the dynamical system. That may happen, for example, in a solid if a continuous mass distribution is taken into account.

If the expansion is carried further, additional corrections might appear. From eqs. (37) through (40) it can be seen that the odd corrections always cancel out. The next term will contain the fourth derivative of the wave function, the second derivative with a coefficient of order R, and the wave function multiplied by a coefficient containing the second covariant derivative of the curvature tensor and contractions of the curvature itself. Thus, the next correction will be altogether of order $\hbar^3\tau R^2$, hence vanishingly small.

[1]B. S, De Witt, Rev. Mod. Phys. 29, 377 (1957).
[2]K. S. Cheng, J. Math. Phys. 13, 1723 (1972).
[3]R. P. Feynman, Rev. Mod. Phys. 20, 327 (1948).
[4]J. A. Schouten, *Ricci-Calculus* (Springer, Berlin, 1954), 2nd ed.
[5]O. Veblen, *Invariants of Quadratic Differential Forms*, (Cambridge U.P., Cambridge, 1962).
[6]In Ref. 2 the factor 2 is missing from the rhs. The error, however, cancels out by repeating it in the normalization factor.

Proof of the strong subadditivity of quantum-mechanical entropy

Elliott H. Lieb[*†]

I.H.E.S., 91 Bures-sur-Yvette, France

Mary Beth Ruskai[*‡]

Department of Mathematics, M.I.T., Cambridge, Massachusetts 02139
(Received 12 April 1973)

We prove several theorems about quantum-mechanical entropy, in particular, that it is strongly subadditive.

1. INTRODUCTION

In this paper we prove several theorems about quantum mechanical entropy, in particular, that it is strongly subadditive (SSA). These theorems were announced in an earlier note,[1] to which we refer the reader for a discussion of the physical significance of SSA and for a review of the historical background. We repeat here a bibliography of relevant papers.[2-9]

The setting for these theorems is as follows:

(a) Given a separable Hilbert space H and a positive, trace-class operator, ρ, on H [i.e., $\rho \geq 0$ means $(\psi, \rho\psi) \geq 0$ for all ψ in H], the entropy of ρ is defined to be

$$S(\rho) \equiv -\operatorname{Tr}\rho \ln\rho = -\sum_{i=1}^{\infty} \lambda_i \ln\lambda_i, \tag{1.1}$$

where Tr means trace, the λ_i are the eigenvalues of ρ, $0 \ln 0 \equiv 0$, and we permit the possibility $S(\rho) = \infty$. In physical applications one also requires that $\operatorname{Tr}\rho = 1$, in which case ρ is called a density matrix.

(b) If $H_{12} = H_1 \otimes H_2$ is the tensor product of two Hilbert spaces and ρ_{12} is a positive, trace-class operator on H_{12}, we can define a positive, trace-class operator, ρ_1, on H_1 by the partial trace, i.e.,

$$\rho_1 \equiv \operatorname{Tr}_2 \rho_{12} \tag{1.2}$$

by which we mean

$$(\varphi, \rho_1 \psi) = \sum_{i=1}^{\infty} (\varphi \otimes e_i, \rho_{12}[\psi \otimes e_i]) \tag{1.3}$$

for all φ, ψ in H_1 and $\{e_i\}_{i=1}^{\infty}$ any orthonormal basis in H_2. We shall denote $S(\rho_1)$ by S_1, etc. In like manner one can have $H_{123} = H_1 \otimes H_2 \otimes H_3$, and ρ_{123} a positive, trace-class operator on H_{123}, and define ρ_{12} on $H_{12} \equiv H_1 \otimes H_2$, ρ_1 on H_1, etc. by partial traces. When no confusion arises, we shall frequently use the symbol ρ_1 to denote the operator $\rho_1 \otimes 1_2$ on H_{12}.

Our main results are the following two theorems.

Theorem 1: Let $H_{12} = H_1 \otimes H_2$. Then the function

$$\rho_{12} \mapsto S_1 - S_{12} \tag{1.4}$$

is convex on the set of positive, trace-class operators on H_{12}.

Theorem 2 (Strong Subadditivity): Let H_{123} and ρ_{123} be defined as in (b) above. Then

(i) $$S_{123} + S_2 - S_{12} - S_{23} \leq 0 \tag{1.5}$$

and

(ii) $$S_1 + S_3 - S_{12} - S_{23} \leq 0. \tag{1.6}$$

In the next section we prove these theorems in the finite-dimensional case. In Sec. 3 we elucidate the connection between these two theorems and give some related results. Sec. 4 contains the proofs for the infinite-dimensional case and is based on the appendix kindly contributed by B. Simon, to whom we are most grateful.

2. PROOFS OF THEOREMS 1 AND 2 IN THE FINITE-DIMENSIONAL CASE

Proof of Theorem 1: The theorem states that

$$(S_1 - S_{12})(\rho_{12}) \leq \alpha(S_1 - S_{12})(\rho'_{12}) + (1-\alpha)(S_1 - S_{12})(\rho''_{12}) \tag{2.1}$$

where $\rho_{12} = \alpha\rho'_{12} + (1-\alpha)\rho''_{12}$, $0 \leq \alpha \leq 1$, and ρ'_{12} and ρ''_{12} are any positive, trace-class operators on H_{12}. We shall assume that both ρ'_{12} and ρ''_{12} are strictly positive and appeal to continuity of $\rho \mapsto S(\rho)$ in the semidefinite case. Letting

$$\Delta = \alpha \operatorname{Tr}_{12}\rho'_{12}(-\ln\rho'_{12} + \ln\rho'_1 + \ln\rho_{12} - \ln\rho_1)$$

and

$$\Gamma = (1-\alpha)\operatorname{Tr}_{12}\rho''_{12}(-\ln\rho''_{12} + \ln\rho''_1 + \ln\rho_{12} - \ln\rho_1),$$

one sees that (2.1) is equivalent to $\Delta + \Gamma \leq 0$. We now use Klein's inequality[7,10]:

$$\operatorname{Tr}(-A\ln A + A\ln B) \leq \operatorname{Tr}(B - A). \tag{2.2}$$

(Alternatively, one could use the Peierls–Bogoliubov inequality in a similar way.[2]) We first apply (2.2) to Δ with $A = \rho'_{12}$ and $B = \exp(\ln\rho'_1 + \ln\rho_{12} - \ln\rho_1)$ and then similarly to Γ. Then

$$\begin{aligned}\Delta + \Gamma &\leq \alpha \operatorname{Tr}_{12}[\exp(\ln\rho'_1 + \ln\rho_{12} - \ln\rho_1) - \rho'_{12}] \\ &\quad + (1-\alpha)\operatorname{Tr}_{12}[\exp(\ln\rho''_1 + \ln\rho_{12} - \ln\rho_1) - \rho''_{12}] \\ &\leq \operatorname{Tr}_{12}[\exp(\ln\rho_1 + \ln\rho_{12} - \ln\rho_1) - \rho_{12}] = 0.\end{aligned} \tag{2.3}$$

The second inequality in (2.3) follows from the concavity[11] of $C \mapsto \operatorname{Tr}[\exp(K + \ln C)]$ for positive C applied to $\rho_1 = \alpha\rho'_1 + (1-\alpha)\rho''_1$ with $K = \ln\rho_{12} - \ln\rho_1$. **Q.E.D.**

Proof of Theorem 2: It has already been pointed out[2] that (1.5) and (1.6) are equivalent; however, we shall prove each statement separately.

(i) Proof of (1.5): We use Klein's inequality, (2.2), with $A = \rho_{123}$ and $B = \exp(-\ln\rho_2 + \ln\rho_{12} + \ln\rho_{23})$. One finds

$$\begin{aligned}F(\rho_{123}) &\equiv S_{123} + S_2 - S_{12} - S_{23} \\ &\leq \operatorname{Tr}_{123}[\exp(\ln\rho_{12} - \ln\rho_2 + \ln\rho_{23}) - \rho_{123}].\end{aligned}$$

We now apply a generalization[11] of the Golden–Thompson inequality, i.e.,

$$\mathrm{Tr}[\exp(\ln B - \ln C + \ln D)] \le \mathrm{Tr}\int_0^\infty B(C+x\mathbb{1})^{-1}D(C+x\mathbb{1})^{-1}dx. \quad (2.4)$$

Thus

$$\begin{aligned} F(\rho_{123}) &\le \mathrm{Tr}_{123}\Big(\int_0^\infty \rho_{12}(\rho_2+x\mathbb{1})^{-1} \\ &\quad\times \rho_{23}(\rho_2+x\mathbb{1})^{-1}dx - \rho_{123}\Big) \\ &= \mathrm{Tr}_2\int_0^\infty \rho_2(\rho_2+x\mathbb{1})^{-1}\rho_2(\rho_2+x\mathbb{1})^{-1}dx - \mathrm{Tr}_{123}\rho_{123} \\ &= \mathrm{Tr}_2\rho_2 - \mathrm{Tr}_{123}\rho_{123} = 0. \end{aligned}$$

Q.E.D.

(ii) Proof of (1.6): Call the left side of (1.6) $G(\rho_{123})$. Note that $S_1 - S_{12}$ is convex in ρ_{12} by Theorem 1; since ρ_{12} is linear in ρ_{123}, $S_1 - S_{12}$ is convex in ρ_{123}. Thus, $G(\rho_{123})$ is convex in ρ_{123}. In the convex cone of positive matrices, the extremal rays consist of matrices of the form $\rho = \alpha P$ where $\alpha \ge 0$ and P is a one-dimensional projection. If ρ_{123} is extremal, then (see Ref. 2, Lemma 3) $S_1 = S_{23}$ and $S_3 = S_{12}$, so that $G(\rho_{123}) = 0$. Every positive matrix ρ_{123} can be written as a convex combination of extremal matrices; it then follows from the convexity of G that $G(\rho_{123}) \le 0$. Q.E.D.

3. REMARKS AND RELATED RESULTS

We have already noted in the proof of (1.6) that Theorem 1 implies Theorem 2. We now note that the converse is also true and give several alternative proofs of Theorems 1 and 2. We then show that $F(\rho_{123})$ is not convex and give a corollary to Theorem 1.

(A) To show Theorem 2 implies Theorem 1 it suffices to note that [apart from the trivial interchange of the subscripts 1 and 2 in (2.1)] (1.5) is identical to (2.1) for a special choice of ρ_{123}, i.e., $\rho_{123} = \alpha\rho'_{12}\otimes E_3 + (1-\alpha)\rho''_{12}\otimes F_3$ where H_3 is chosen to be two-dimensional and E_3 and F_3 are orthogonal, one-dimensional projections on H_3.

(B) Uhlmann[9] has shown that (1.5) follows from the concavity of $C \mapsto \mathrm{Tr}\exp(K + \ln C)$. This has been shown to be true by Lieb,[11] and an alternate proof was later found by Epstein.[12] Therefore, Uhlmann's remark gives an alternate proof of (1.5).

(C) The proof of (1.6) shows that Theorem 1 implies Theorem 2. However, (1.6) is not equivalent to (1.5) in other contexts.[13] [In fact, (1.6) is false in the classical continuous case.[6]] Therefore, it is instructive to note that one can show that Theorem 1 implies (1.5) directly without using (1.6). Baumann and Jost[3,5] have shown that a special choice of ρ'_{12} and ρ''_{12} in (2.1) implies that $\mathrm{Tr}\int_0^\infty A^*(C+x\mathbb{1})^{-1}A(C+x\mathbb{1})^{-1}dx$ is jointly convex in (A, C) where A and C are matrices with $C > 0$. Lieb has then shown[11] that this implies $C \mapsto \mathrm{Tr}\exp(K + \ln C)$ is concave in C. The last statement was used to prove[11] (2.4) which, as we have already seen, implies (1.5). Alternatively, we have already noted in (B) above that concavity of $C \mapsto \mathrm{Tr}\exp[K + \ln C]$ implies (1.5).

(D) We have already shown that the left side of (1.6), $G(\rho_{123})$, is convex. One might wonder, therefore, if the left side of (1.5), $F(\rho_{123})$, is also convex. In fact, it is not. If it were, one could choose H_2 to be one-dimensional so that

$$F(\rho_{123}) = S_{13} - S_1 - S_3 \equiv E(\rho_{13})$$

would have to be a convex function of ρ_{13}. Take H_1 and H_3 to be two-dimensional and choose ρ'_{13} and ρ''_{13} to be the following orthogonal, one-dimensional projections:

$$\rho'_{13}(i_1,i_3;j_1,j_3) = \tfrac{1}{2}\delta(i_1,i_3)\delta(j_1,j_3)$$

and

$$\rho''_{13}(i_1,i_3;j_1,j_3) = \tfrac{1}{2}[1-\delta(i_1,i_3)][1-\delta(j_1,j_3)],$$

where δ is the Krönecker delta. Then $\rho'_1 = \rho''_1 = \frac{1}{2}\mathbb{1}_1$, $\rho'_3 = \rho''_3 = \frac{1}{2}\mathbb{1}_3$, and $E(\rho'_{13}) + E(\rho''_{13}) - 2E(\frac{1}{2}\rho'_{13} + \frac{1}{2}\rho''_{13}) = -2\ln 2 < 0$, which is a contradiction.

(E) It was pointed out in Ref. 11 that if $f(A)$ is a convex function from the set of positive matrices into $\mathbb{R}$, and if it is also homogeneous [i.e., $f(\lambda A) = \lambda f(A)$ for all $\lambda > 0$], then

$$\frac{d}{dx}f(A+xB)\Big|_{x=0} \equiv \lim_{x\downarrow 0} x^{-1}[f(A+xB) - f(A)] \le f(B), \quad (3.1)$$

whenever A, B are positive matrices and the above limit exists. The function $(S_1 - S_{12})(\rho_{12})$ has these properties. To apply (3.1) we compute

$$\begin{aligned} \frac{d}{dx}S(\rho+x\gamma) &= -\frac{d}{dx}\mathrm{Tr}[(\rho+x\gamma)\ln(\rho+x\gamma)] \\ &= -\mathrm{Tr}\gamma\ln(\rho+x\gamma) - \mathrm{Tr}\gamma. \end{aligned}$$

Using this in (3.1) we conclude

Corollary: Let γ_{12} and ρ_{12} be positive, trace-class matrices on H_{12}. Then

$$\mathrm{Tr}_{12}\gamma_{12}\ln\rho_{12} - \mathrm{Tr}_1\gamma_1\ln\rho_1 \le \mathrm{Tr}_{12}\gamma_{12}\ln\gamma_{12} - \mathrm{Tr}_1\gamma_1\ln\gamma_1, \quad (3.2)$$

i.e., for each fixed γ_{12}, the left side of (3.2) achieves its maximum when $\rho_{12} = \gamma_{12}$.

4. EXTENSION TO INFINITE-DIMENSIONS

We can use Theorem A2 to extend Theorems 1 and 2 to infinite dimensions. For simplicity, we confine our discussion to Theorem 1 where $H_{12} = H_1 \otimes H_2$. The extension of Theorem 2 is similar and we point out the necessary changes at the end of this section.

Let E_i^n $(i = 1, 2$ and $n = 1, 2, \cdots)$ be sequences of increasing, finite-dimensional projections on H_i, converging strongly to the identity, and define

$$E^n = E_1^n \otimes E_2^n,$$

$$\rho_{12}^n = E^n\rho_{12}E^n,$$

and

$$\rho_1^n = \mathrm{Tr}_2\rho_{12}^n = E_1^n(\mathrm{Tr}_2 E_2^n\rho_{12}E_2^n)E_1^n \quad (4.1)$$

Since the spaces $E_i^n H_i$ are finite dimensional, Theorem 1 is satisfied by ρ_{12}^n on $E_1^nH_1 \otimes E_2^nH_2$ for each n. Thus, it suffices to show that the sequences of matrices $\{\rho_{12}^n\}_{n=1}^\infty$ and $\{\rho_1^n\}_{n=1}^\infty$ satisfy the hypotheses of Theorem A2 so that, e.g., $\lim_{n\to\infty} S(\rho_{12}^n) = S(\rho_{12}) = S_{12}$.

To show that $\{\rho_{12}^n\}_{n=1}^\infty$ satisfies Theorem A2, we first note that $E^n \xrightarrow{s} \mathbb{1}_{12}$. If[14] the sequences $A_n \xrightarrow{s} A$ and $B_n \xrightarrow{s} B$, then $A_nB_n \xrightarrow{s} AB$. Consequently, ρ_{12}^n converges to ρ_{12} strongly, and therefore weakly. It follows from the Ritz principle (see Proposition A1) that $\rho_{12}^n = E^n\rho_{12}E^n \lhd E^{n+1}\rho_{12}E^{n+1} \lhd \rho_{12}$, with $\lhd$ as defined

in the Appendix. Therefore, the hypotheses of Theorem A2 are satisfied and

$$\lim_{n\to\infty} S(\rho_{12}^n) = S_{12}. \tag{4.2}$$

To show that $\{\rho_1^n\}_{n=1}^\infty$ also satisfies Theorem A. 2, define $\tilde{\rho}_1^n = \mathrm{Tr}_2 E_2^n \rho_{12} E_2^n$. Then $\rho_1^n = E_1^n \tilde{\rho}_1^n E_1^n$. To show that ρ_1^n converges to ρ_1 weakly, it suffices to show that $\tilde{\rho}_1^n$ converges to ρ_1^n strongly. (In fact, it converges uniformly.) To do this we can assume, without loss of generality, that E_2^n projects on the space spanned by $e_i \cdots e_n$ where $\{e_i : i = 1, 2, \cdots\}$ is an orthonormal basis in H_2. Then

$$(\psi, \tilde{\rho}_1^n \psi) = \sum_{i=1}^n (\psi \otimes e_i, \rho_{12}\psi \otimes e_i)$$

for all ψ in H_1, and it follows that

$$\tilde{\rho}_1^n \le \tilde{\rho}_1^{n+1}, \tag{4.3}$$

and

$$\lim_{n\to\infty} (\psi, (\rho_1 - \tilde{\rho}_1^n)\psi) = \lim_{n\to\infty} \sum_{n+1}^\infty (\psi \otimes e_i, \rho_{12}\psi \otimes e_i) = 0 \tag{4.4}$$

Since $\tilde{\rho}_1^n$ is a monotone sequence of positive operators, (4.4) implies that $\tilde{\rho}_1^n \xrightarrow{s} \rho_1$ and therefore $\rho_1^n \xrightarrow{s} \rho_1$. Further, it follows from (4.3), i.e., the monotonicity of $\tilde{\rho}_1^n$, that

$$\begin{aligned}\rho_1^n &\lhd E_1^{n+1} \tilde{\rho}_1^n E_1^{n+1} \\ &\le E_1^{n+1} \tilde{\rho}_1^{n+1} E_1^{n+1} = \rho_1^{n+1} \lhd \rho_1.\end{aligned}$$

Thus, Theorem A2 implies

$$\lim_{n\to\infty} S(\rho_1^n) = S(\rho_1) = S_1.$$

The analysis for Theorem 2 is similar. One defines

$$E^n = E_1^n \otimes E_2^n \otimes E_3^n,$$

$$\rho_{123}^n = E^n \rho_{123} E^n,$$

and

$$\rho_{12}^n = \mathrm{Tr}_3 \rho_{123}^n, \text{etc.}$$

ACKNOWLEDGMENTS

We thank D. Ruelle for generating and encouraging our interest in this problem. We are also grateful to R. Jost, O. Lanford, and D. Robinson for their encouragement and for helpful conversations. One author (E.L.) would like to thank the Chemistry Laboratory III, University of Copenhagen, where part of this work was done.

APPENDIX: CONVERGENCE THEOREMS FOR ENTROPY By B. Simon §

We discuss a variety of convergence theorems which are useful in extending entropy inequalities from finite dimensional matrices to infinite dimensional operators on a Hilbert space.

Definition: Let A be a positive compact operator. $\mu_k(A)$ denotes the kth largest eigenvalue of A counting multiplicity.

Definition: Let $s(x)$ be the function on $[0, \infty)$ given by

$$s(x) = \begin{cases} -x \ln x & \text{if } x \ge 0 \\ 0 & \text{if } x = 0. \end{cases}$$

If A is positive and compact, we set

$$S(A) = \sum_{k=1}^\infty s(\mu_k(A)),$$

the value infinity being allowed.

Definition: Let A and B be positive, compact operators. We write $A \lhd B$ if and only if $\mu_k(A) \le \mu_k(B)$ for all k.

Definition: Let $\{A_n\}_{n=1}^\infty$ and A be positive, compact operators. We write $A_n \xrightarrow{\mu} A$ if and only if $\mu_k(A_n) \longrightarrow \mu_k(A)$ for each fixed k.

Remarks: (1) The topology defined by μ-convergence is, of course, non-Hausdorff. (2) The order $\lhd$ is useful because of the following consequence of the Ritz principle:

Proposition A1: Let A be a positive, compact operator and let P be a projection. Then $PAP \lhd A$. In particular, if P and Q are projections and $P \le Q$, then $PAP \lhd QAQ$.

The above is false if $\lhd$ is replaced by $\le$.

Theorem A1 (Basic Convergence Theorem): Let B be a positive, compact operator with $S(B) < \infty$. Suppose $\{A_n\}$ and A are given positive, compact operators with

(1) $A_n \xrightarrow{\mu} A$,

(2) $A_n \lhd B$ for each n.

Then, $\lim_{n\to\infty} S(A_n) = S(A)$.

Proof: The proof is based on the fact that s is monotone in $[0, e^{-1}]$. Since B is compact, $\mu_k(B) \longrightarrow 0$. Suppose $\mu_N(B) \le e^{-1}$. By (1) and the continuity of s, $s(\mu_k(A_n)) \longrightarrow s(\mu_k(A))$, each k, and by (2) and the monotonicity of s in $[0, e^{-1}]$, $s(\mu_k(A_n)) \le s(\mu_k(B))$ for $k \ge N$, each n. Thus by the dominated convergence theorem for sums, $\sum_{k\ge N} s(\mu_k(A_n)) \longrightarrow \sum_{k\ge N} s(\mu_k(A))$. Since $\sum_{k\le N-1} s(\mu_k(A_n))$ certainly converges, the theorem is proven. Q.E.D.

For applications of Theorem A1, it is convenient to have statements expressed in a more usual form than μ-convergence.

Theorem A2: Let $\{A_n\}$ and A be positive, compact operators. If

(1) $\underset{n\to\infty}{\text{w-lim}}\, A_n = A$

and

(2) $A_n \lhd A$ for all n,

then $\lim_{n\to\infty} S(A_n) = S(A)$.

Proof: We first prove that $A_n \xrightarrow{\mu} A$. Fix k and ϵ. By weak convergence and the min–max principle, it is easy to find a k-dimensional space, V, and an N such that

$$(\psi, A_n \psi) \ge (\mu_k(A) - \epsilon)\|\psi\|^2$$

if $\psi \in V$ and $n \ge N$. But then $\mu_k(A_n) \ge \mu_k(A) - \epsilon$ if $n \ge N$. Since $\mu_k(A) \ge \mu_k(A_n)$ by (2), this means $|\mu_k(A) - \mu_k(A_n)| < \epsilon$ if $n \ge N$ and hence $A_n \xrightarrow{\mu} A$. If $S(A) < \infty$, the theorem then follows from Theorem A1. If $S(A) = \infty$, for any M we can find an L such that

$\sum_{k=1}^{L} s(\mu_k(A)) > M$. However, for L sufficiently large, $S(A_n) \geq \sum_{k=1}^{L} s(\mu_k(A_n))$ and, since $\mu_k(A_n) \longrightarrow \mu_k(A)$, the latter sum can be made arbitrarily close to M. Thus $S(A_n) \longrightarrow \infty$. Q.E.D.

Theorem A3: (Dominated Convergence Theorem for Entropy): Let $\{A_n\}, A$ and B be positive, compact operators and suppose that

(1) $S(B) < \infty$,

(2) $\underset{n\to\infty}{\text{w-lim}}\, A_n = A$,

(3) $A_n \leq B$ (operator inequality!).

Then,

$$\lim_{n\to\infty} S(A_n) = S(A).$$

Proof: Since B is compact, for any $\epsilon > 0$ we can find a finite-dimensional subspace $K \subset H$ such that $(u, Bu) = \|B^{1/2}u\| < \epsilon\|u\|$ for $u \in L$, where L is the orthogonal complement of K. Since $A_n \leq B$, $\|A_n^{1/2}u\| = (u, A_n u) \leq (u, Bu) \leq \epsilon\|u\|$ for all u in L. Since $A_n \xrightarrow{w} A$, $A \leq B$, and $\|A^{1/2}u\| \leq \epsilon\|u\|$ for all u in L also. We now show $A_n \longrightarrow A$ uniformly. Recall that $\|A_n - A\| = \text{supp}\{|(\varphi, (A_n - A)\psi)| : \varphi, \psi \in H, \|\varphi\| = \|\psi\| = 1\}$. Now write $\varphi = f + u$, $\psi = g + v$ where f, g are in K and u, v in L. Then

$$\begin{aligned}(\varphi, (A_n - A)\psi) &= ((f+u), (A_n - A)(g+v)) \\ &\leq (f, (A_n - A)g) + \|A_n^{1/2}f\|^{1/2}\|A_n^{1/2}v\|^{1/2} \\ &\quad + \|A^{1/2}f\|^{1/2}\|A^{1/2}v\|^{1/2} + \|A_n^{1/2}u\|^{1/2}\|A_n^{1/2}g\|^{1/2} \\ &\quad + \|A^{1/2}u\|^{1/2}\|A^{1/2}g\|^{1/2} + \|A_n^{1/2}u\|^{1/2}\|A_n^{1/2}v\|^{1/2} \\ &\quad + \|A^{1/2}u\|^{1/2}\|A^{1/2}v\|^{1/2},\end{aligned}$$

which can be arbitrarily small since $A_n \longrightarrow A$ uniformly on K, $A_n^{1/2}$ and $A^{1/2}$ are bounded on K, $\|A_n^{1/2}u\| < \epsilon$, $\|A^{1/2}u\| < \epsilon$, etc., and $\|f\| \leq \|\varphi\|$, etc. Thus $|(\varphi, (A_n - A)\psi)|$ can be made arbitrarily small independent of φ, ψ (for all φ, ψ with $\|\varphi\| = \|\psi\| = 1$) and thus $\|A_n - A\| \longrightarrow 0$. By the min-max principle, $|\mu_k(A_n) - \mu_k(A)| \leq \|A_n - A\|$. Thus $A_n \xrightarrow{\mu} A$, and (1) implies that Theorem A1 is applicable. Q.E.D.

Example: Let $\{A_n\}, A$ and B be the following operators on H, where $\{\varphi_n\}$ is an orthonormal basis for H:

$$A\varphi_k = 0, \quad \text{each } k,$$
$$A_n\varphi_k = \delta_{nk}e^{-1}\varphi_n,$$
$$B = A_1.$$

Then $A_n \not\lhd B$, $A_n \longrightarrow A$ strongly, but $S(A_n)$ does not converge to $S(A)$. This example shows that $\leq$ and not $\not\lhd$ is needed in Theorem A3.

*Work partially supported by U.S. National Science Foundation Grant GP-31674 X.

†On leave from Department of Mathematics, M.I.T., Cambridge, Mass. 02139, U.S.A. Work partially supported by a Guggenheim Memorial Foundation fellowship.

‡Supported in part by the National Research Council of Canada Grant No. NRC-A6595 at the University of Alberta, Edmonton, Canada. Present address: Department of Mathematics, University of Oregon, Eugene, Oregon 97403.

§Princeton University; A. Sloan Fellow.

[1] E. H. Lieb and M. B. Ruskai, Phys. Rev. Letters **30**, 434 (1973).
[2] H. Araki and E. H. Lieb, Commun. Math. Phys. **18**, 160 (1970).
[3] F. Bauman and R. Jost, in *Problems of Theoretical Physics; Essays Dedicated to N. N. Bogoliubov* (Moscow, Nauka, 1969), p. 285.
[4] R. Jost, in *Quanta: Essays in Theoretical Physics Dedicated to Gregor Wentzel,* edited by P. G. O. Freund, C. J. Goebel and Y. Nambu (University of Chicago Press, Chicago, 1970), p. 13.
[5] F. Baumann, Helv. Phys. Acta **44**, 95 (1971).
[6] D. W. Robinson and D. Ruelle, Commun. Math. Phys. **5**, 288 (1967).
[7] O. Lanford III and D. W. Robinson, J. Math. Phys. **9**, 1120 (1968).
[8] E. P. Wigner and M. M. Yanase, Proc. Nat. Acad. Sci. **49**, 910 (1963); Can. J. Math. **16**, 397 (1964).
[9] A. Uhlmann, "Endlich Dimensionale Dichtematrizen, II". Wiss. Z. Karl-Marx-University Leipzig, Math-Naturwiss. R. **22**, Jg. H. 2, 139 (1973).
[10] D. Ruelle, *Statistical Mechanics: Rigorous Results* (Benjamin, New York, 1969), Theorem 2.5.2.
[11] E. H. Lieb, "Convex Trace Functions and the Wigner-Yanase-Dyson Conjecture", Adv. in Math., to appear Dec. 1973.
[12] H. Epstein, Commun. Math. Phys. **37**, 317 (1973).
[13] M. B. Ruskai, "A Generalization of the Entropy Using Traces on von Neumann Algebras," preprint.
[14] O. Lanford III, in *Statistical Mechanics and Quantum Field Theory* edited by C. De Witt and R. Stora (Gordon and Breach, New York, 1971), p. 174.

Statistical theory of effective electrical, thermal, and magnetic properties of random heterogeneous materials. II. Bounds for the effective permittivity of statistically anisotropic materials

Motoo Hori

Department of Applied Physics, Tokyo Institute of Technology, Meguroku, Tokyo 152, Japan
(Received 22 May 1973)

Upper and lower bounds are derived for the effective permittivity of random heterogeneous materials that are statistically homogeneous but not necessarily statistically isotropic. As trial functions in standard variational principles, we use modified perturbation expansions for the electric field and the electric displacement. The bounds are expressed in terms of the many-point correlation functions of the spatial variation of permittivity. For a wide class of random multiphase materials called cell materials, explicit calculation is performed taking account of the three-point correlation effects.

1. INTRODUCTION

In a previous paper,[1] hereafter referred to as I, a general perturbation formulation was developed for the effective permittivity of random heterogeneous materials which are statistically homogeneous but not always statistically isotropic. Especially, the second-order and third-order perturbation terms were explicitly determined on the basis of the modified cell model. The purpose of the present paper is to derive upper and lower bounds on the effective permittivity of such inhomogeneous anisotropic media, by utilizing as trial functions the perturbation series obtained in I.

As for bounding effective physical constants of statistically isotropic materials, a considerable amount of work has been done since the pioneering research of Wiener.[2] Hashin and Shtrikman[3,4] improved Wiener's bounds with the aid of some variational principles. Their results were expressed in terms of the one-point moments of the permittivity field or the volume fractions of the constituting phases. Brown,[5] Beran,[6] and Beran and Molyneux[7] gave more restrictive bounds which involved such additional statistical information as three-point correlation functions. Miller[8,9] and Beran and Silnutzer[10] showed that for the so-called cell materials the effects of third-order correlations may be represented by means of the shape factors of cells. In this article, the variational techniques developed by Beran[6] will be extended to treat the statistically anisotropic case.

2. BASIC CONCEPTS AND VARIATIONAL PRINCIPLES

Let us consider a heterogeneous material of volume V and surface S whose local permittivity ϵ can be regarded as a random function of position $\mathbf{r}$. The material volume V is taken to be infinite. Suppose that the medium is statistically homogeneous, that is to say all correlation functions of $\epsilon(\mathbf{r})$ are independent of absolute positions. Then, the n-point moment $\langle\epsilon(\mathbf{r}_1)\epsilon(\mathbf{r}_2)\cdots\epsilon(\mathbf{r}_n)\rangle$ becomes

$$\langle\epsilon(\mathbf{r}_1)\epsilon(\mathbf{r}_2)\cdots\epsilon(\mathbf{r}_n)\rangle = \langle\epsilon(0)\epsilon(\mathbf{r}_{12})\cdots\epsilon(\mathbf{r}_{1n})\rangle, \tag{2.1}$$

where the brackets $\langle\ \rangle$ indicate the ensemble average and $\mathbf{r}_{ij}$ designate the relative positions $\mathbf{r}_j - \mathbf{r}_i$. In connection with the assumption of statistical homogeneity, we make an ergodic hypothesis such that the volume average equals the ensemble average in the limit as $V \to \infty$.

We shall confine ourselves to the case where a constant-average electric field is applied to a statistically anisotropic medium. The effective permittivity tensor ϵ_{ij}^* is defined by the relation

$$\langle D_i\rangle = \langle\epsilon(\mathbf{r})E_i(\mathbf{r})\rangle = \epsilon_{ij}^*\langle E_j\rangle. \tag{2.2}$$

Here E_i and D_i are the ith components of the electric field and the electric displacement, respectively, and the summation convention has been employed. Since ϵ_{ij}^* is a symmetric tensor of rank two, there exist three principal values $\epsilon_1^*, \epsilon_2^*, \epsilon_3^*$. When the coordinate axes are chosen to coincide with the principal axes, we have

$$\langle D_i\rangle = \epsilon_{(i)}^*\langle E_i\rangle, \tag{2.3}$$

the parentheses around the index i implying no summation on i.

The ensemble average of the electrostatic energy density for the material with ϵ_{ij}^* may be written as

$$\langle u\rangle = \tfrac{1}{2}\langle\epsilon(\mathbf{r})E_i(\mathbf{r})E_i(\mathbf{r})\rangle = \tfrac{1}{2}\epsilon_{ij}^*\langle E_i\rangle\langle E_j\rangle. \tag{2.4}$$

Alternatively, if we denote the inverse matrix of ϵ_{ij}^* by κ_{ij}^*,

$$\langle u\rangle = \tfrac{1}{2}\langle D_i(\mathbf{r})D_i(\mathbf{r})/\epsilon(\mathbf{r})\rangle = \tfrac{1}{2}\kappa_{ij}^*\langle D_i\rangle\langle D_j\rangle. \tag{2.5}$$

Along the principal axes Eqs. (2.4) and (2.5) reduce to

$$\langle u\rangle = \tfrac{1}{2}\epsilon_{(i)}^*\langle E_i\rangle\langle E_i\rangle = \tfrac{1}{2}\langle E_i\rangle\langle E_i\rangle/\kappa_{(i)}^*, \tag{2.6}$$

$$\langle u\rangle = \tfrac{1}{2}\kappa_{(i)}^*\langle D_i\rangle\langle D_i\rangle = \tfrac{1}{2}\langle D_i\rangle\langle D_i\rangle/\epsilon_{(i)}^*. \tag{2.7}$$

The ergodic hypothesis enables us to recast Eqs. (2.4) and (2.5) as

$$\langle u\rangle = \frac{1}{2V}\int_V \epsilon(\mathbf{r})E_i(\mathbf{r})E_i(\mathbf{r})d\omega, \tag{2.8}$$

$$\langle u\rangle = \frac{1}{2V}\int_V \frac{D_i(\mathbf{r})D_i(\mathbf{r})}{\epsilon(\mathbf{r})}d\omega, \tag{2.9}$$

where $d\omega$ represents a volume element of V at $\mathbf{r}$.

The standard variational principles established by Beran[6,11] are as follows:

Principle 1: Let $E_i(\mathbf{r})$ be a statistically homogeneous vector field. The integral

$$\langle\epsilon(\mathbf{r})E_i(\mathbf{r})E_i(\mathbf{r})\rangle = \frac{1}{V}\int_V \epsilon(\mathbf{r})E_i(\mathbf{r})E_i(\mathbf{r})d\omega \tag{2.10}$$

subject to the subsidiary condition

$$E_i(\mathbf{r}) = \frac{\partial \Phi(\mathbf{r})}{\partial x_i} \tag{2.11}$$

is stationary and a minimum for

$$\frac{\partial}{\partial x_i}[\epsilon(\mathbf{r})E_i(\mathbf{r})] = 0. \tag{2.12}$$

In Eq. (2.11) $\Phi(\mathbf{r})$ is a scalar field; instead of a boundary condition we impose the requirement that any trial function $E_i(\mathbf{r})$ must have the same average as the true function.

Principle 2: Let $D_i(\mathbf{r})$ be a statistically homogeneous vector field. The integral

$$\left\langle \frac{D_i(\mathbf{r})D_i(\mathbf{r})}{\epsilon(\mathbf{r})} \right\rangle = \frac{1}{V}\int_V \frac{D_i(\mathbf{r})D_i(\mathbf{r})}{\epsilon(\mathbf{r})}\,d\omega \tag{2.13}$$

subject to the subsidiary condition

$$D_i(\mathbf{r}) = e_{ijk}\frac{\partial A_k(\mathbf{r})}{\partial x_j} \tag{2.14}$$

is stationary and a minimum for

$$e_{ijk}\frac{\partial}{\partial x_j}\frac{D_k(\mathbf{r})}{\epsilon(\mathbf{r})} = 0. \tag{2.15}$$

In Eqs. (2.14) and (2.15), $A_k(\mathbf{r})$ is a vector field and e_{ijk} is the third-order alternating tensor of Levi–Civita ($e_{ijk} = 1$ or -1 according as i,j,k is an even or odd permutation of 1, 2, 3; otherwise $e_{ijk} = 0$). Instead of a boundary condition we impose the requirement that any trial function $D_i(\mathbf{r})$ must have the same average as the true function.

The simplest assumption we can make is to take the correct averages as trial functions. By using the mean electric field $\langle E_i\rangle$ as an admissible field for Principle 1, we obtain

$$(\epsilon_{ij}^* - \langle\epsilon\rangle)\langle E_i\rangle\langle E_j\rangle \le 0. \tag{2.16}$$

The inequality (2.16) is equivalent to

$$(\epsilon_{(i)}^* - \langle\epsilon\rangle)\langle E_i\rangle\langle E_i\rangle \le 0, \tag{2.17}$$

provided that the principal axes are parallel to the coordinate axes. In order that the quadratic form on the left be nonpositive definite, it is necessary and sufficient that

$$\epsilon_i^* \le \langle\epsilon\rangle. \tag{2.18}$$

Similarly, choosing the mean electric displacement $\langle D_i\rangle$ as an admissible field for Principle 2, we get

$$\kappa_i^* = 1/\epsilon_i^* \le \langle 1/\epsilon\rangle. \tag{2.19}$$

Combination of the above two inequalities yields

$$\langle 1/\epsilon\rangle^{-1} \le \epsilon_i^* \le \langle\epsilon\rangle, \tag{2.20}$$

which is an extension of Wiener's elementary bounds to heterogeneous media with statistical anisotropy.

3. DERIVATION OF UPPER BOUNDS

According to the perturbation theory described in I, the electric field in a statistically homogeneous and anisotropic material can be expanded in a series of the form

$$E_i(\mathbf{r}) = \langle E_i\rangle + E_i^{(1)}(\mathbf{r}) + E_i^{(2)}(\mathbf{r}) + \cdots. \tag{3.1}$$

The nth-order perturbation term $E_i^{(n)}(\mathbf{r})$ is given by

$$E_i^{(n)}(\mathbf{r}) = \frac{\langle E_j\rangle}{(4\pi\langle\epsilon\rangle)^n}\int_V d\omega_{12}\int_V d\omega_{23}\cdots\int_V d\omega_{n,n+1} \times \frac{x_{12,i}}{r_{12}^3}\frac{x_{23,k}}{r_{23}^3}\cdots\frac{\partial^n \epsilon'(\mathbf{r}_2)\epsilon'(\mathbf{r}_3)\cdots\epsilon'(\mathbf{r}_{n+1})}{\partial x_{12,k}\partial x_{23,h}\cdots\partial x_{n,n+1;j}}, \tag{3.2}$$

where $\epsilon'(\mathbf{r}) = \epsilon(\mathbf{r}) - \langle\epsilon\rangle$. Needless to say, $E_i^{(n)}(\mathbf{r})$ is of the same order of smallness as $[\epsilon'(\mathbf{r})/\langle\epsilon\rangle]^n$ and satisfies $\langle E_i^{(n)}(\mathbf{r})\rangle = 0$. Convergence of the series (3.1) is assumed but not proven if $\langle|\epsilon'^n|\rangle/\langle\epsilon\rangle^n \ll 1$.

Similarly, it was pointed out in I that the effective permittivity tensor ϵ_{ij}^* may also be expressed as

$$\epsilon_{ij}^* = \langle\epsilon\rangle\left(\delta_{ij} - \sum_{n=2}^{\infty}(-1)^n A_{ij}^{(n)}\frac{\langle\epsilon'^n\rangle}{\langle\epsilon\rangle^n}\right), \tag{3.3}$$

δ_{ij} denoting the Kronecker delta. The nth-order perturbation coefficient $A_{ij}^{(n)}$ is related to the normalized n-point correlation function of $\epsilon(\mathbf{r})$,

$$f(\mathbf{r}_{12},\mathbf{r}_{13},\ldots,\mathbf{r}_{1n}) = \frac{\langle\epsilon'(\mathbf{r}_1)\epsilon'(\mathbf{r}_2)\cdots\epsilon'(\mathbf{r}_n)\rangle}{\langle\epsilon'^n\rangle}, \tag{3.4}$$

by

$$A_{ij}^{(n)} = \left(\frac{-1}{4\pi}\right)^{n-1}\int_V d\omega_{12}\int_V d\omega_{23}\cdots\int_V d\omega_{n-1,n} \times\frac{x_{12,i}}{r_{12}^3}\frac{x_{23,k}}{r_{23}^3}\cdots\frac{\partial^{n-1}f(\mathbf{r}_{12},\mathbf{r}_{13},\cdots,\mathbf{r}_{1n})}{\partial x_{12,k}\partial x_{23,h}\cdots\partial x_{n-1,n;j}}. \tag{3.5}$$

Furthermore, we postulate that all of the tensors $A_{ij}^{(n)}$ must possess common principal axes. In fact, for a cell material composed of uniformly oriented ellipsoidal cells, it can be shown that the principal axes of $A_{ij}^{(2)}$ and $A_{ij}^{(3)}$ coincide with the axes of the ellipsoids.

As an admissible solution for use in Principle 1, we will adopt a modification of the perturbation expansion (3.1) that is obtained by the inclusion of a set of multiplicative constants. Namely, we intend to introduce the trial function of the form

$$E_{N,i}(\mathbf{r}) = \langle E_i\rangle + \lambda_1 E_i^{(1)}(\mathbf{r}) + \lambda_2 E_i^{(2)}(\mathbf{r}) + \cdots + \lambda_N E_i^{(N)}(\mathbf{r}), \tag{3.6}$$

which obeys the condition $\langle E_{N,i}(\mathbf{r})\rangle = \langle E_i\rangle$. The multiplicative constants λ_n are to be chosen to minimize the upper bound on the effective permittivity, when this sum of a finite number of terms is used as an admissible solution. If an infinite number of terms were considered, we would have $\lambda_n = 1$. In the case of $N = 0$, we of course, find the elementary bound presented in Eq. (2.18).

Substitution of the admissible field $E_{N,i}(\mathbf{r})$ into the variational integral yields

$$\begin{aligned}\epsilon_{ij}^*\langle E_i\rangle\langle E_j\rangle &\le \langle\epsilon(\mathbf{r})E_{N,i}(\mathbf{r})E_{N,i}(\mathbf{r})\rangle \\ &= \langle\epsilon\rangle\langle E_i\rangle^2 + 2\langle E_i\rangle\sum_{n=1}^{N}\lambda_n\langle\epsilon' E_i^{(n)}\rangle \\ &\quad + \langle\epsilon\rangle\sum_{n=1}^{N}\sum_{m=1}^{N}\lambda_n\lambda_m\langle E_i^{(n)}E_i^{(m)}\rangle \\ &\quad + \sum_{n=1}^{N}\sum_{m=1}^{N}\lambda_n\lambda_m\langle\epsilon' E_i^{(n)}E_i^{(m)}\rangle.\end{aligned} \tag{3.7}$$

In Appendix A we show that under appropriate boundary conditions

$$\langle\epsilon' E_i^{(n)}\rangle = (-1)^n A_{ij}^{(n+1)}(\langle\epsilon'^{n+1}\rangle/\langle\epsilon\rangle^n)\langle E_j\rangle, \tag{3.8}$$

$$\langle E_i^{(n)} E_i^{(m)} \rangle = (-1)^{n+m} A_{jk}^{(n+m)} (\langle \epsilon'^{n+m} \rangle / \langle \epsilon \rangle^{n+m}) \langle E_j \rangle \langle E_k \rangle, \tag{3.9}$$

$$\langle \epsilon' E_i^{(n)} E_i^{(m)} \rangle = (-1)^{n+m} A_{jk}^{(n+m+1)} (\langle \epsilon'^{n+m+1} \rangle / \langle \epsilon \rangle^{n+m}) \langle E_j \rangle \langle E_k \rangle, \tag{3.10}$$

whence

$$\Bigg[\epsilon_{ij}^* - \langle \epsilon \rangle \Bigg(\delta_{ij} + 2 \sum_1^N (-1)^n \lambda_n A_{ij}^{(n+1)} \frac{\langle \epsilon'^{n+1} \rangle}{\langle \epsilon \rangle^{n+1}} + \sum_1^N \sum_1^N (-1)^{n+m} \lambda_n \lambda_m A_{ij}^{(n+m)} \frac{\langle \epsilon'^{n+m} \rangle}{\langle \epsilon \rangle^{n+m}} + \sum_1^N \sum_1^N (-1)^{n+m} \lambda_n \lambda_m A_{ij}^{(n+m+1)} \frac{\langle \epsilon'^{n+m+1} \rangle}{\langle \epsilon \rangle^{n+m+1}} \Bigg) \Bigg] \langle E_i \rangle \langle E_j \rangle \leq 0. \tag{3.11}$$

In view of the theory of quadratic forms, it follows from the above inequality that for ϵ_i^*

$$\epsilon_i^* \leq \langle \epsilon \rangle \Bigg(1 + 2 \sum_1^N (-1)^n \lambda_n A_i^{(n+1)} \frac{\langle \epsilon'^{n+1} \rangle}{\langle \epsilon \rangle^{n+1}} + \sum_1^N \sum_1^N (-1)^{n+m} \lambda_n \lambda_m A_i^{(n+m)} \frac{\langle \epsilon'^{n+m} \rangle}{\langle \epsilon \rangle^{n+m}} + \sum_1^N \sum_1^N (-1)^{n+m} \lambda_n \lambda_m A_i^{(n+m+1)} \frac{\langle \epsilon'^{n+m+1} \rangle}{\langle \epsilon \rangle^{n+m+1}} \Bigg), \tag{3.12}$$

where $A_i^{(n)}$ indicate the eigenvalues of the symmetric second-order tensors $A_{ij}^{(n)}$. For later convenience, we rewrite Eq. (3.12) as

$$\epsilon_i^* \leq \langle \epsilon \rangle \Bigg(1 - \sum_{n=2}^{N+1} (-1)^n \lambda_{n-1} (2 - \lambda_1) A_i^{(n)} \frac{\langle \epsilon'^n \rangle}{\langle \epsilon \rangle^n} - \sum_{n=N+2}^{2N+1} (-1)^n \lambda_{n-N-1} \lambda_N A_i^{(n)} \frac{\langle \epsilon'^n \rangle}{\langle \epsilon \rangle^n} + \sum_{n=1}^N \sum_{m=2}^N (-1)^{n+m} \lambda_n (\lambda_m - \lambda_{m-1}) A_i^{(n+m)} \frac{\langle \epsilon'^{n+m} \rangle}{\langle \epsilon \rangle^{n+m}} \Bigg). \tag{3.13}$$

We first consider the case when $N = 1$. Setting $\lambda_n = 0$ for $n \geq 2$, we have

$$\epsilon_i^* \leq \langle \epsilon \rangle [1 - \lambda_1 (2 - \lambda_1) A_i^{(2)} \langle \epsilon'^2 \rangle / \langle \epsilon \rangle^2 + \lambda_1^2 A_i^{(3)} \langle \epsilon'^3 \rangle / \langle \epsilon \rangle^3]. \tag{3.14}$$

In order to obtain the best upper bound for ϵ_i^*, the right-hand side of Eq. (3.14) must be minimized with respect to λ_1. By the way, Eqs. (3.9) and (3.10) lead to

$$\langle \epsilon E_i^{(1)} E_i^{(1)} \rangle = \langle \epsilon \rangle \langle E_i^{(1)} E_i^{(1)} \rangle + \langle \epsilon' E_i^{(1)} E_i^{(1)} \rangle = \langle \epsilon \rangle \left(A_{jk}^{(2)} \frac{\langle \epsilon'^2 \rangle}{\langle \epsilon \rangle^2} + A_{jk}^{(3)} \frac{\langle \epsilon'^3 \rangle}{\langle \epsilon \rangle^3} \right) \langle E_j \rangle \langle E_k \rangle \geq 0, \tag{3.15}$$

so that

$$A_i^{(2)} \langle \epsilon'^2 \rangle / \langle \epsilon \rangle^2 + A_i^{(3)} \langle \epsilon'^3 \rangle / \langle \epsilon \rangle^3 \geq 0. \tag{3.16}$$

Therefore, the expression on the right-hand side of Eq. (3.14) is found to have a minimum. As a final result we obtain

$$\epsilon_i^* \leq \langle \epsilon \rangle \left(1 - \frac{(A_i^{(2)} \langle \epsilon'^2 \rangle / \langle \epsilon \rangle^2)^2}{A_i^{(2)} \langle \epsilon'^2 \rangle / \langle \epsilon \rangle^2 + A_i^{(3)} \langle \epsilon'^3 \rangle / \langle \epsilon \rangle^3} \right). \tag{3.17}$$

When the medium is statistically isotropic, the tensors ϵ_{ij}^* and $A_{ij}^{(n)}$ reduce to scalars ϵ^* and $A^{(n)}$, respectively. Thus, the upper bound in Eq. (3.17) becomes

$$\epsilon^* \leq \langle \epsilon \rangle \left(1 - \frac{\frac{1}{9} \langle \epsilon'^2 \rangle^2 / \langle \epsilon \rangle^4}{\frac{1}{3} \langle \epsilon'^2 \rangle / \langle \epsilon \rangle^2 + A^{(3)} \langle \epsilon'^3 \rangle / \langle \epsilon \rangle^3} \right), \tag{3.18}$$

considering that $A^{(2)} = \frac{1}{3}$. The third-order perturbation coefficient $A^{(3)}$ is defined as

$$A^{(3)} = \frac{1}{(4\pi)^2} \int_V d\omega_{21} \int_V d\omega_{23} \frac{x_{21,i}}{r_{21}^3} \frac{x_{23,k}}{r_{23}^3} \frac{\partial^2 f(\mathbf{r}_{21}, \mathbf{r}_{23})}{\partial x_{21,k} \partial x_{23,(i)}}, \tag{3.19}$$

with

$$f(\mathbf{r}_{21}, \mathbf{r}_{23}) = \langle \epsilon'(\mathbf{r}_1) \epsilon'(\mathbf{r}_2) \epsilon'(\mathbf{r}_3) \rangle / \langle \epsilon'^3 \rangle = \langle \epsilon'(0) \epsilon'(\mathbf{r}_{21}) \epsilon'(\mathbf{r}_{23}) \rangle / \langle \epsilon'^3 \rangle. \tag{3.20}$$

Under boundary conditions as stated in I [see Eqs. (2.57) and (2.58) in I], Eq. (3.19) may be transformed to

$$A^{(3)} = \frac{1}{(4\pi)^2} \int_V d\omega_{21} \int_V d\omega_{23} \frac{x_{21,k}}{r_{21}^3} \frac{x_{23,k}}{r_{23}^3} \frac{\partial^2 f(\mathbf{r}_{21}, r_{23})}{\partial x_{21,i} \partial x_{23,(i)}}. \tag{3.21}$$

This means that the upper bound (3.18) is in agreement with the result reached by Beran.[6]

Next, let us discuss the case of $N = 2$. Then Eq. (3.13) reduces to

$$\epsilon_i^* \leq \langle \epsilon \rangle [1 - \lambda_1 (2 - \lambda_1) A_i^{(2)} \langle \epsilon'^2 \rangle / \langle \epsilon \rangle^2 + (\lambda_1^2 - 2\lambda_1 \lambda_2 + 2\lambda_2) A_i^{(3)} \langle \epsilon'^3 \rangle / \langle \epsilon \rangle^3 - \lambda_2 (2\lambda_1 - \lambda_2) A_i^{(4)} \langle \epsilon'^4 \rangle / \langle \epsilon \rangle^4 + \lambda_2^2 A_i^{(5)} \langle \epsilon'^5 \rangle / \langle \epsilon \rangle^5]. \tag{3.22}$$

We notice that there exists a minimum of the right-hand side of the inequality (3.22), provided that

$$(A_{(i)}^{(2)} \langle \epsilon'^2 \rangle / \langle \epsilon \rangle^2 + A_{(i)}^{(3)} \langle \epsilon'^3 \rangle / \langle \epsilon \rangle^3)(A_i^{(4)} \langle \epsilon'^4 \rangle / \langle \epsilon \rangle^4 + A_i^{(5)} \langle \epsilon'^5 \rangle / \langle \epsilon \rangle^5) \geq (A_{(i)}^{(3)} \langle \epsilon'^3 \rangle / \langle \epsilon \rangle^3 + A_{(i)}^{(4)} \langle \epsilon'^4 \rangle / \langle \epsilon \rangle^4)^2. \tag{3.23}$$

The above requirement is certainly fulfilled; for one has

$$\langle \epsilon E_{(i)}^{(1)} E_{(i)}^{(1)} \rangle \langle \epsilon E_{(i)}^{(2)} E_i^{(2)} \rangle \geq \langle \epsilon E_{(i)}^{(1)} E_{(i)}^{(2)} \rangle^2, \tag{3.24}$$

which follows directly from Schwarz' inequality. The best bound is given by

$$\epsilon_i^* \leq \langle \epsilon \rangle \left(1 - \frac{F(\langle \epsilon'^2 \rangle / \langle \epsilon \rangle^2, \langle \epsilon'^3 \rangle / \langle \epsilon \rangle^3, \langle \epsilon'^4 \rangle / \langle \epsilon \rangle^4, \langle \epsilon'^5 \rangle / \langle \epsilon \rangle^5)}{G(\langle \epsilon'^2 \rangle / \langle \epsilon \rangle^2, \langle \epsilon'^3 \rangle / \langle \epsilon \rangle^3, \langle \epsilon'^4 \rangle / \langle \epsilon \rangle^4, \langle \epsilon'^5 \rangle / \langle \epsilon \rangle^5)} \right), \tag{3.25}$$

where

$$F(\xi_2, \xi_3, \xi_4, \xi_5) = (A_{(i)}^{(2)} \xi_2)^2 (A_i^{(4)} \xi_4 + A_i^{(5)} \xi_5) - 2 A_{(i)}^{(2)} \xi_2 \cdot A_{(i)}^{(3)} \xi_3 (A_i^{(3)} \xi_3 + A_i^{(4)} \xi_4) + (A_{(i)}^{(3)} \xi_3)^2 (A_i^{(2)} \xi_2 + A_i^{(3)} \xi_3), \tag{3.26}$$

$$G(\xi_2, \xi_3, \xi_4, \xi_5) = (A_{(i)}^{(2)} \xi_2 + A_{(i)}^{(3)} \xi_3)(A_i^{(4)} \xi_4 + A_i^{(5)} \xi_5) - (A_{(i)}^{(3)} \xi_3 + A_{(i)}^{(4)} \xi_4)^2. \tag{3.27}$$

It is extremely difficult to derive the best bound for arbitrary N. Hence we make a simplifying assumption that $\lambda_1 = \lambda_2 = \cdots = \lambda_N = 1$, although it does not produce the best bound. From Eq. (3.13) we easily find

$$\epsilon_i^* \leq \langle \epsilon \rangle \left(1 - \sum_{n=2}^{2N+1} (-1)^n A_i^{(n)} \frac{\langle \epsilon'^n \rangle}{\langle \epsilon \rangle^n} \right). \tag{3.28}$$

The right-hand side is nothing but the perturbation expansion of ϵ_i^* up to $(2N + 1)$th order. It is worth while to note that a finite perturbation series terminated at odd order provides an upper bound. This result corresponds with the nonnegative property of even-order

perturbation coefficients $A_i^{(2m)}$, which is an immediate consequence of Eq. (3.9).

4. DERIVATION OF LOWER BOUNDS

Proceeding in the same way as we did in the foregoing section, we shall seek a lower bound for ϵ_i^*. By the analogy of Eq. (3.1) we write

$$D_i(\mathbf{r}) = \langle D_i\rangle + D_i^{(1)}(\mathbf{r}) + D_i^{(2)}(\mathbf{r}) + \cdots, \tag{4.1}$$

where $D_i^{(n)}(\mathbf{r})$ is of order $[\epsilon'(\mathbf{r})/\langle\epsilon\rangle]^n$ and satisfies $\langle D_i^{(n)}(\mathbf{r})\rangle = 0$. The relationship of $E_i^{(n)}(\mathbf{r})$ and $D_i^{(n)}(\mathbf{r})$ can be expressed by

$$\begin{aligned} D_i^{(n)}(\mathbf{r}) &= \langle\epsilon\rangle E_i^{(n)}(\mathbf{r}) + \epsilon'(\mathbf{r})E_i^{(n-1)}(\mathbf{r}) - \langle\epsilon' E_i^{(n-1)}\rangle \\ &= \langle\epsilon\rangle H_{ij}^{(n)}(\mathbf{r})\langle E_j\rangle. \end{aligned} \tag{4.2}$$

Equation (4.2) applies even to the case $n = 1$, if we interpret $D_i^{(0)}(\mathbf{r})$ as $\langle D_i\rangle$. The trial function introduced for $D_i(\mathbf{r})$ is

$$D_{N,i}(\mathbf{r}) = \langle D_i\rangle + \mu_1 D_i^{(1)}(\mathbf{r}) + \mu_2 D_i^{(2)}(\mathbf{r}) + \cdots + \mu_N D_i^{(N)}(\mathbf{r}), \tag{4.3}$$

which includes adjustable constants μ_n and meets the condition $\langle D_{N,i}(\mathbf{r})\rangle = \langle D_i\rangle$.

Making use of Principle 2, we have

$$\begin{aligned} \epsilon_{ij}^*\langle E_i\rangle\langle E_j\rangle &\le \left\langle \frac{D_{N,i}(\mathbf{r})D_{N,i}(\mathbf{r})}{\epsilon(\mathbf{r})}\right\rangle \\ &= \left\langle\frac{1}{\epsilon}\right\rangle \langle D_i\rangle^2 + 2\langle D_i\rangle\sum_{n=1}^{N}\mu_n\left\langle\frac{D_i^{(n)}}{\epsilon}\right\rangle \\ &\quad + \sum_{n=1}^{N}\sum_{m=1}^{N}\mu_n\mu_m\left\langle\frac{D_i^{(n)}D_i^{(m)}}{\epsilon}\right\rangle. \end{aligned} \tag{4.4}$$

Insertion of Eq. (4.2) into Eq. (4.4) gives

$$\begin{aligned} \epsilon_{ij}^*\langle E_i\rangle\langle E_j\rangle \le \Bigg(&\left\langle\frac{1}{\epsilon}\right\rangle\epsilon_{ij}^*\epsilon_{ik}^* + 2\langle\epsilon\rangle\epsilon_{ij}^*\sum_1^N\mu_n\left\langle\frac{H_{ik}^{(n)}}{\epsilon}\right\rangle \\ &+ \langle\epsilon\rangle^2\sum_1^N\sum_1^N\mu_n\mu_m\left\langle\frac{H_{ij}^{(n)}H_{ik}^{(m)}}{\epsilon}\right\rangle\Bigg)\langle E_j\rangle\langle E_k\rangle. \end{aligned} \tag{4.5}$$

When the coordinate axes are taken to be parallel with the principal axes of ϵ_{ij}^*, it holds that

$$\begin{aligned} \kappa_i^* = \frac{1}{\epsilon_i^*} \le \left\langle\frac{1}{\epsilon}\right\rangle &+ 2\langle\epsilon\rangle\sum_1^N\frac{\mu_n}{\epsilon_i^*}\left\langle\frac{H_i^{(n)}}{\epsilon}\right\rangle \\ &+ \langle\epsilon\rangle^2\sum_1^N\sum_1^N\frac{\mu_n}{\epsilon_{(i)}^*}\frac{\mu_m}{\epsilon_{(i)}^*}\left\langle\frac{H_{(i)}^{(n)}H_i^{(m)}}{\epsilon}\right\rangle. \end{aligned} \tag{4.6}$$

This is just a fundamental inequality for determining a lower bound on ϵ_i^*. Putting $N = 0$ in Eq. (4.6), we arrive at the elementary bound given by Eq. (2.19).

Now let us define a new type of normalized n-point correlation function as

$$\begin{aligned} &g_l(\mathbf{r}_{12},\mathbf{r}_{13},\ldots,\mathbf{r}_{1n}) \\ &\quad = \langle\epsilon'(\mathbf{r}_1)\epsilon'(\mathbf{r}_2)\cdots\epsilon'(\mathbf{r}_{l-1})\kappa'(\mathbf{r}_l)\epsilon'(\mathbf{r}_{l+1}) \\ &\qquad \cdots\epsilon'(\mathbf{r}_n)\rangle/\langle\kappa'\epsilon'^{n-1}\rangle, \end{aligned} \tag{4.7}$$

where $\kappa'(\mathbf{r}) = \kappa(\mathbf{r}) - \langle\kappa\rangle = 1/\epsilon(\mathbf{r}) - \langle 1/\epsilon\rangle$. The tensor quantity corresponding to $A_{ij}^{(n)}$ is

$$B_{ij}^{(n,l)} = \left(\frac{-1}{4\pi}\right)^{n-1}\int_V d\omega_{12}\int_V d\omega_{23}\cdots\int_V d\omega_{n-1,n} \times\frac{x_{12,i}}{r_{12}^3}\frac{x_{23,k}}{r_{23}^3}\cdots\frac{\partial^{n-1}g_l(\mathbf{r}_{12},\mathbf{r}_{13},\cdots,\mathbf{r}_{1n})}{\partial x_{12,k}\partial x_{23,h}\cdots\partial x_{n-1,n;j}}. \tag{4.8}$$

As demonstrated in Appendix B, it turns out that under usual boundary conditions

$$\langle\kappa' E_i^{(n)}\rangle = (-1)^n B_{ij}^{(n+1,1)}(\langle\kappa'\epsilon'^n\rangle/\langle\epsilon\rangle^n)\langle E_j\rangle, \tag{4.9}$$

$$\langle\kappa' E_i^{(n)}E_i^{(m)}\rangle = (-1)^{n+m}B_{jk}^{(n+m+1,n+1)} \times(\langle\kappa'\epsilon'^{n+m}\rangle/\langle\epsilon\rangle^{n+m})\langle E_j\rangle\langle E_k\rangle. \tag{4.10}$$

On the other hand we can deduce a recurrence formula

$$\frac{\langle\kappa'\epsilon'^n\rangle}{\langle\kappa\rangle\langle\epsilon\rangle^n} = -\left[\frac{\langle\kappa'\epsilon'^{n-1}\rangle}{\langle\kappa\rangle\langle\epsilon\rangle^{n-1}} + \frac{\langle\epsilon'^n\rangle}{\langle\epsilon\rangle^n} + \left(1 - \frac{1}{\langle\kappa\rangle\langle\epsilon\rangle}\right)\frac{\langle\epsilon'^{n-1}\rangle}{\langle\epsilon\rangle^{n-1}}\right]. \tag{4.11}$$

For $n = 1$ we get

$$\langle\kappa'\epsilon'\rangle/\langle\kappa\rangle\langle\epsilon\rangle = -(1 - 1/\langle\kappa\rangle\langle\epsilon\rangle). \tag{4.12}$$

Repeated application of Eq. (4.11) yields

$$\begin{aligned} \frac{\langle\kappa'\epsilon'^n\rangle}{\langle\kappa\rangle\langle\epsilon\rangle^n} = &-\frac{\langle\epsilon'^n\rangle}{\langle\epsilon\rangle^n} - \frac{1}{\langle\kappa\rangle\langle\epsilon\rangle}\sum_{l=2}^{n-1}(-1)^{n+l}\frac{\langle\epsilon'^l\rangle}{\langle\epsilon\rangle^l} \\ &+ (-1)^n\left(1 - \frac{1}{\langle\kappa\rangle\langle\epsilon\rangle}\right). \end{aligned} \tag{4.13}$$

Consequently, $\langle H_i^{(n)}/\epsilon\rangle$ and $\langle H_{(i)}^{(n)}H_i^{(m)}/\epsilon\rangle$ appearing in Eq. (4.6) are represented as functions of $A_i^{(2)}, A_i^{(3)}, \ldots, B_i^{(2,1)}, B_i^{(3,1)}, \ldots, \langle\epsilon\rangle, \langle\kappa\rangle, \langle\epsilon'^2\rangle, \langle\epsilon'^3\rangle, \cdots$.

We shall again begin with the case $N = 1$. In this case Eq. (4.6) is simplified to

$$\begin{aligned} \frac{\kappa_i^*}{\langle\kappa\rangle} \le 1 &- 2\frac{\langle\epsilon\rangle}{\epsilon_i^*}\mu_1(1 - B_i^{(2,1)})\left(1 - \frac{1}{\langle\kappa\rangle\langle\epsilon\rangle}\right) + \left(\frac{\langle\epsilon\rangle}{\epsilon_i^*}\mu_1\right)^2 \\ &\times\left[(1 - 2B_i^{(2,1)} + B_i^{(3,2)})\left(1 - \frac{1}{\langle\kappa\rangle\langle\epsilon\rangle}\right)\right. \\ &\left. + (A_i^{(2)} - B_i^{(3,2)})\frac{\langle\epsilon'^2\rangle}{\langle\epsilon\rangle^2}\right], \end{aligned} \tag{4.14}$$

since

$$\begin{aligned} \left\langle\frac{H_i^{(1)}}{\epsilon}\right\rangle &= \langle\kappa\rangle(1 - B_i^{(2,1)})\frac{\langle\kappa'\epsilon'\rangle}{\langle\kappa\rangle\langle\epsilon\rangle} \\ &= -\langle\kappa\rangle(1 - B_i^{(2,1)})\left(1 - \frac{1}{\langle\kappa\rangle\langle\epsilon\rangle}\right), \end{aligned} \tag{4.15}$$

$$\begin{aligned} &\left\langle\frac{H_{(i)}^{(1)}H_i^{(1)}}{\epsilon}\right\rangle \\ &\quad = \langle\kappa\rangle\left[A_i^{(2)}\frac{\langle\epsilon'^2\rangle}{\langle\epsilon\rangle^2} + (2B_i^{(2,1)} - 1)\frac{\langle\kappa'\epsilon'\rangle}{\langle\kappa\rangle\langle\epsilon\rangle}\right. \\ &\qquad \left. + B_i^{(3,2)}\frac{\langle\kappa'\epsilon'^2\rangle}{\langle\kappa\rangle\langle\epsilon\rangle^2}\right] \\ &\quad = \langle\kappa\rangle\left[(1 - 2B_i^{(2,1)} + B_i^{(3,2)})\left(1 - \frac{1}{\langle\kappa\rangle\langle\epsilon\rangle}\right)\right. \\ &\qquad \left. + (A_i^{(2)} - B_i^{(3,2)})\frac{\langle\epsilon'^2\rangle}{\langle\epsilon\rangle^2}\right]. \end{aligned} \tag{4.16}$$

Maximizing the right-hand side of Eq. (4.14) with respect to μ_1, we obtain

$$\epsilon_i^* \geq \frac{1}{\langle 1/\epsilon\rangle}\Big\{1-(1-B_i^{(2,1)})^2\Big(1-\frac{1}{\langle\kappa\rangle\langle\epsilon\rangle}\Big)^2 \times\Big[(1-2B_i^{(2,1)}+B_i^{(3,2)})\Big(1-\frac{1}{\langle\kappa\rangle\langle\epsilon\rangle}\Big) + (A_i^{(2)}-B_i^{(3,2)})\frac{\langle\epsilon'^2\rangle}{\langle\epsilon\rangle^2}\Big]^{-1}\Big\}^{-1}. \tag{4.17}$$

For statistically isotropic materials we observe that

$$B_{ij}^{(2,1)} = -\frac{1}{4\pi\langle\kappa'\epsilon'\rangle}\int_V d\omega_{12}\frac{x_{12,i}}{r_{12}^3}\frac{\partial\langle\kappa'(\mathbf{r}_1)\epsilon'(\mathbf{r}_2)\rangle}{\partial x_{12,j}} = B^{(2,1)}\delta_{ij} = \tfrac{1}{3}\delta_{ij}, \tag{4.18}$$

$$B_{ij}^{(3,2)} = \frac{1}{(4\pi)^2\langle\kappa'\epsilon'^2\rangle}\int_V d\omega_{21}\int_V d\omega_{23}\frac{x_{21,i}}{r_{21}^3}\frac{x_{23,k}}{r_{23}^3} \times\frac{\partial^2\langle\epsilon'(\mathbf{r}_1)\kappa'(\mathbf{r}_2)\epsilon'(\mathbf{r}_3)\rangle}{\partial x_{21,k}\partial x_{23,j}} = B^{(3,2)}\delta_{ij}. \tag{4.19}$$

Equation (4.18) can be established exactly as the formula $A^{(2)} = \frac{1}{3}$ given in Sec. 2C of I. Accordingly,

$$\epsilon^* \geq \frac{1}{\langle 1/\epsilon\rangle}\{1-\tfrac{4}{3}(1-1/\langle\kappa\rangle\langle\epsilon\rangle)^2 \times[(\tfrac{1}{3}+B_i^{(3,2)})(1-1/\langle\kappa\rangle\langle\epsilon\rangle) + (\tfrac{1}{3}-B_i^{(3,2)})\langle\epsilon'^2\rangle/\langle\epsilon\rangle^2]^{-1}\}^{-1}. \tag{4.20}$$

It is readily confirmed that this bound agrees with the result of Beran.[6]

The expressions for the best lower bound are much complicated when $N \geq 2$. Therefore, in lieu of attempting to derive the best bound for arbitrary N, we want to get a less intricate bound by supposing $\mu_1 = \mu_2 = \cdots = \mu_N = \epsilon_i^*/\langle\epsilon\rangle$. Straightforward but rather tedious calculation shows

$$\frac{\kappa_i^*}{\langle\kappa\rangle} \leq 1+\frac{\langle\kappa'\epsilon'\rangle}{\langle\kappa\rangle\langle\epsilon\rangle}+\frac{1}{\langle\kappa\rangle\langle\epsilon\rangle}\sum_2^{2N-1}(-1)^n A_i^{(n)}\frac{\langle\epsilon'^n\rangle}{\langle\epsilon\rangle^n} + A_i^{(2N)}\frac{\langle\epsilon'^{2N}\rangle}{\langle\epsilon\rangle^{2N}}+\Big(\sum_2^N(-1)^n A_i^{(n)}\frac{\langle\epsilon'^n\rangle}{\langle\epsilon\rangle^n}\Big)^2 + 2(-1)^N B_i^{(N+1,1)}\frac{\langle\kappa'\epsilon'^N\rangle}{\langle\kappa\rangle\langle\epsilon\rangle^N}\sum_2^N(-1)^n A_i^{(n)}\frac{\langle\epsilon'^n\rangle}{\langle\epsilon\rangle^n} + B_i^{(2N+1,N+1)}\frac{\langle\kappa'\epsilon'^{2N}\rangle}{\langle\kappa\rangle\langle\epsilon\rangle^{2N}}, \tag{4.21}$$

which, together with Eq. (3.28), constitutes a useful expression for effective permittivity bounds involving the $(2N+1)$-point correlation functions.

5. EXPLICIT CALCULATION FOR CELL MATERIALS

Miller[8,9] has proposed a cell model to characterize the geometry of random multiphase materials. As explained by him, the geometry of a cell material is defined as any division of the material space into cells which fulfills the following requirements:

(i) The space is completely covered by nonoverlapping cells.

(ii) Cells are distributed in a manner such that the material is statistically homogeneous.

(iii) The material property ϵ of a cell is statistically independent of the material property of any other cell.

Furthermore, the symmetric cell material signifies a cell material satisfying the additional requirement:

(iv) The conditional probabilities of n points being and n' points not being in the same cell of a particular material, given that one point is in a cell of that material, are the same for each material.

In the latter half of I, we have explicitly evaluated two low-order perturbation coefficients, $A_{ij}^{(2)}$ and $A_{ij}^{(3)}$, for symmetric cell materials. In such a symmetric case the independence assumption (iii) asserts that

$$f(\mathbf{r}_{12}) = \langle\epsilon'(\mathbf{r}_1)\epsilon'(\mathbf{r}_2)\rangle/\langle\epsilon'^2\rangle = P(\mathbf{r}_1,\mathbf{r}_2), \tag{5.1}$$

where $P(\mathbf{r}_1,\mathbf{r}_2)$ stands for the probability that both points $\mathbf{r}_1$ and $\mathbf{r}_2$ fall into the same cell. The second-order perturbation coefficient $A_{ij}^{(2)}$ is therefore

$$A_{ij}^{(2)} = -\frac{1}{4\pi}\int_V d\omega_{12}\frac{x_{12,i}}{r_{12}^3}\frac{\partial P(\mathbf{r}_1,\mathbf{r}_2)}{\partial x_{12,j}}. \tag{5.2}$$

Likewise, if we denote by $P(\mathbf{r}_1,\mathbf{r}_2,\mathbf{r}_3)$ the probability that all the points $\mathbf{r}_1,\mathbf{r}_2,\mathbf{r}_3$ lie in the same cell,

$$A_{ij}^{(3)} = \frac{1}{(4\pi)^2}\int_V d\omega_{12}\int_V d\omega_{23}\frac{x_{12,i}}{r_{12}^3}\frac{x_{23,k}}{r_{23}^3}\frac{\partial^2 P(\mathbf{r}_1,\mathbf{r}_2,\mathbf{r}_3)}{\partial x_{12,k}\partial x_{23,j}}. \tag{5.3}$$

Let us first treat a symmetric cell material composed of cells of uniform shape, size, and orientation. Then $A_{ij}^{(2)}$ is equal to the so-called magnetometric demagnetization tensor[12,13] of the cell. In terms of the point-function demagnetization tensor $L_{ij}(\mathbf{r})$, [13] we can write

$$A_{ij}^{(2)} = M_{ij}^{(2)} \equiv \frac{1}{v}\int_v d\omega L_{ij}(\mathbf{r}), \tag{5.4}$$

$$A_{ij}^{(3)} = M_{ij}^{(3)} \equiv \frac{1}{v}\int_v d\omega L_{ik}(\mathbf{r})L_{kj}(\mathbf{r}), \tag{5.5}$$

v indicating the cell volume. It holds also for $B_{ij}^{(2,1)}$ and $B_{ij}^{(3,2)}$ that

$$B_{ij}^{(2,1)} = M_{ij}^{(2)}, \qquad B_{ij}^{(3,2)} = M_{ij}^{(3)}. \tag{5.6}$$

We remark that $M_{ij}^{(3)}$ as well as $M_{ij}^{(2)}$ depends only on the shape of cells and not on their size. This suggests that the bounding equations (3.17) and (4.17) are uniquely determined by the cell shape, even when the medium consists of cells of varying size.

Secondly, we shall deal with the statistically isotropic case where geometrically similar cells are oriented at random. In this case it is seen that

$$A^{(2)} = B^{(2,1)} = \tfrac{1}{3}M_{ii}^{(2)} = \tfrac{1}{3}, \tag{5.7}$$

$$A^{(3)} = B^{(3,2)} = \tfrac{1}{3}M_{ii}^{(3)} = \frac{1}{3v}\int_v d\omega L_{ik}^2(\mathbf{r}). \tag{5.8}$$

In I we proved the inequality $M_{ii}^{(3)} \geq \frac{1}{3}$, while in Appendix C we show that $M_{ii}^{(3)} \leq 1$. To sum up, $A^{(3)}$ is bounded by

$$\tfrac{1}{9} \leq A^{(3)} \leq \tfrac{1}{3}, \tag{5.9}$$

that has already been found by Miller.[8] For ellipsoidal cells the point-function demagnetization tensor $L_{ij}(\mathbf{r})$ is constant throughout the cell volume, so that $A^{(3)} = \frac{1}{3}L_{ij}^2$. In particular, we may assign $A^{(3)} = \frac{1}{3}$ to the shape of plates, $A^{(3)} = \frac{1}{9}$ to the spherical shape, and $A^{(3)} = \frac{1}{6}$ to the shape of rods of circular cross section.

Higher-order perturbation coefficients which enter in Eq. (3.25), (3.28), or (4.21) are much more difficult to compute than the second-order or third-order perturbation coefficient. Take $A_{ij}^{(4)}$ as an example. The four-point moment $\langle\epsilon'(\mathbf{r}_1)\epsilon'(\mathbf{r}_2)\epsilon'(\mathbf{r}_3)\epsilon'(\mathbf{r}_4)\rangle$ assumes $\langle\epsilon'^4\rangle$ when four points are in the same cell, $\langle\epsilon'^2\rangle^2$ when two pairs of points are in two different cells, and zero otherwise; that is,

$$f(\mathbf{r}_{12},\mathbf{r}_{13},\mathbf{r}_{14}) = P(\mathbf{r}_1,\mathbf{r}_2,\mathbf{r}_3,\mathbf{r}_4) + \langle\epsilon'^2\rangle^2/\langle\epsilon'^4\rangle \times [P(\mathbf{r}_1,\mathbf{r}_2;\mathbf{r}_3,\mathbf{r}_4) + P(\mathbf{r}_1,\mathbf{r}_4;\mathbf{r}_2,\mathbf{r}_3) + P(\mathbf{r}_1,\mathbf{r}_3;\mathbf{r}_2,\mathbf{r}_4)]. \quad (5.10)$$

Here $P(\mathbf{r}_1,\mathbf{r}_2,\mathbf{r}_3,\mathbf{r}_4)$ represents the probability that the points $\mathbf{r}_1,\mathbf{r}_2,\mathbf{r}_3,\mathbf{r}_4$ are in the same cell, $P(\mathbf{r}_1,\mathbf{r}_2;\mathbf{r}_3,\mathbf{r}_4)$ the probability that the two pairs of points, $(\mathbf{r}_1,\mathbf{r}_2)$ and $(\mathbf{r}_3,\mathbf{r}_4)$, are in two different cells, and so on.

Accordingly, $A_{ij}^{(4)}$ may be separated into four parts as

$$A_{ij}^{(4)} = A_{1,ij}^{(4)} + (\langle\epsilon'^2\rangle^2/\langle\epsilon'^4\rangle)(A_{2,ij}^{(4)} + A_{3,ij}^{(4)} + A_{4,ij}^{(4)}), \quad (5.11)$$

where

$$A_{1,ij}^{(4)} = -\frac{1}{(4\pi)^3}\int_V d\omega_{12}\int_V d\omega_{23}\int_V d\omega_{34}\frac{x_{12,i}}{r_{12}^3}\frac{x_{23,k}}{r_{23}^3}\frac{x_{34,h}}{r_{34}^3} \times \frac{\partial^3 P(\mathbf{r}_1,\mathbf{r}_2,\mathbf{r}_3,\mathbf{r}_4)}{\partial x_{12,k}\partial x_{23,h}\partial x_{34,j}}, \quad (5.12)$$

$$A_{2,ij}^{(4)} = -\frac{1}{(4\pi)^3}\int_V d\omega_{12}\int_V d\omega_{23}\int_V d\omega_{34}\frac{x_{12,i}}{r_{12}^3}\frac{x_{23,k}}{r_{23}^3}\frac{x_{34,h}}{r_{34}^3} \times \frac{\partial^3 P(\mathbf{r}_1,\mathbf{r}_2;\mathbf{r}_3,\mathbf{r}_4)}{\partial x_{12,k}\partial x_{23,h}\partial x_{34,j}}, \quad (5.13)$$

$$A_{3,ij}^{(4)} = -\frac{1}{(4\pi)^3}\int_V d\omega_{12}\int_V d\omega_{23}\int_V d\omega_{34}\frac{x_{12,i}}{r_{12}^3}\frac{x_{23,k}}{r_{23}^3}\frac{x_{34,h}}{r_{34}^3} \times \frac{\partial^3 P(\mathbf{r}_1,\mathbf{r}_4;\mathbf{r}_2,\mathbf{r}_3)}{\partial x_{12,k}\partial x_{23,h}\partial x_{34,j}}, \quad (5.14)$$

$$A_{4,ij}^{(4)} = -\frac{1}{(4\pi)^3}\int_V d\omega_{12}\int_V d\omega_{23}\int_V d\omega_{34}\frac{x_{12,i}}{r_{12}^3}\frac{x_{23,k}}{r_{23}^3}\frac{x_{34,h}}{r_{34}^3} \times \frac{\partial^3 P(\mathbf{r}_1,\mathbf{r}_3;\mathbf{r}_2,\mathbf{r}_4)}{\partial x_{12,k}\partial x_{23,h}\partial x_{34,j}}. \quad (5.15)$$

Analogously to $A_{ij}^{(2)}$ and $A_{ij}^{(3)}$, $A_{1,ij}^{(4)}$ is independent not only of the size of cells but also of their relative arrangement, because $P(\mathbf{r}_1,\mathbf{r}_2,\mathbf{r}_3,\mathbf{r}_4)$ concerns the geometry of a single cell. On the contrary, the quantities like $P(\mathbf{r}_1,\mathbf{r}_2;\mathbf{r}_3,\mathbf{r}_4)$ refer to the mutual relation between different cells, so that $A_{2,ij}^{(4)}$, $A_{3,ij}^{(4)}$, and $A_{4,ij}^{(4)}$ are essentially influenced by the frequency distribution of dimensions or distances of cells. In this sense, the expressions (3.17) and (4.17) taking into account the three-point correlations provide the best possible bounds in terms of the cell shape alone. These bounds are

$$\epsilon_i^* \le \langle\epsilon\rangle\left(1 - \frac{(M_{(i)}^{(2)}\langle\epsilon'^2\rangle/\langle\epsilon\rangle^2)^2}{M_i^{(2)}\langle\epsilon'^2\rangle/\langle\epsilon\rangle^2 + M_i^{(3)}\langle\epsilon'^3\rangle/\langle\epsilon\rangle^3}\right), \quad (5.16)$$

$$\epsilon_i^* \ge (1/\langle 1/\epsilon\rangle)\{1 - (1 - M_{(i)}^{(2)})^2(1 - 1/\langle\kappa\rangle\langle\epsilon\rangle)^2 \times [(1 - 2M_i^{(2)} + M_i^{(3)})(1 - 1/\langle\kappa\rangle\langle\epsilon\rangle) + (M_i^{(2)} - M_i^{(3)})\langle\epsilon'^2\rangle/\langle\epsilon\rangle^2]^{-1}\}^{-1}. \quad (5.17)$$

Finally we are concerned with an asymmetric cell material in which the statistical properties of the geometry of cells of different phases are dissimilar. Let v_l and $A_{ij}^{(n)}(l)$ be the volume fraction and the nth-order perturbation coefficient of the lth phase with property ϵ_l, respectively. As discussed in I, the values of $A_{ij}^{(2)}$ and $A_{ij}^{(3)}$ for an asymmetric cell material can be estimated from those for a symmetric cell material; namely,

$$A_{ij}^{(2)} = (1/\langle\epsilon'^2\rangle)\sum_l v_l(\epsilon_l - \langle\epsilon\rangle)^2 A_{ij}^{(2)}(l), \quad (5.18)$$

$$A_{ij}^{(3)} = (1/\langle\epsilon'^3\rangle)\sum_l v_l(\epsilon_l - \langle\epsilon\rangle)^3 A_{ij}^{(3)}(l). \quad (5.19)$$

It should be noted that such a simple averaging formula does not hold for $n \ge 4$.

ACKNOWLEDGMENTS

The author wishes to express his sincere thanks to Mr. Y. Fujiyama and Mr. T. Okada for their assistance during the course of this work. He is also grateful to the referee who made many helpful comments on the earlier paper and the present paper.

APPENDIX A: EVALUATION OF $\langle\epsilon' E_i^{(n)}\rangle$, $\langle E_i^{(n)}E_i^{(m)}\rangle$, AND $\langle\epsilon' E_i^{(n)}E_i^{(m)}\rangle$

We shall present a detailed proof of Eqs. (3.8)–(3.10). In the first place, Eq. (3.8) is obvious from the definitions of $E_i^{(n)}(\mathbf{r})$ and $A_{ij}^{(n)}$.

In the next place, let us seek $\langle E_i^{(n)}E_i^{(m)}\rangle$ and prove Eq. (3.9). By way of illustration consider the case of $n = 1$ and $m = 2$. It follows from Eq. (3.2) that

$$\langle E_i^{(1)}(\mathbf{r}_1)E_i^{(2)}(\mathbf{r}_1)\rangle = \frac{\langle E_j\rangle\langle E_k\rangle}{(4\pi\langle\epsilon\rangle)^3}\int_V d\omega_{12}\int_V d\omega_{13}\int_V d\omega_{34} \times \frac{x_{12,i}}{r_{12}^3}\frac{x_{13,i}}{r_{23}^3}\frac{x_{34,k}}{r_{34}^3}\frac{\partial^3\langle\epsilon'(\mathbf{r}_2)\epsilon'(\mathbf{r}_3)\epsilon'(\mathbf{r}_4)\rangle}{\partial x_{12,j}\partial x_{13,h}\partial x_{34,k}}. \quad (A1)$$

Transformation of the variables of integration from $\mathbf{r}_{12},\mathbf{r}_{13},\mathbf{r}_{34}$ to $\mathbf{r}_{12},\mathbf{r}_{23} = \mathbf{r}_{13} - \mathbf{r}_{12},\mathbf{r}_{34}$ gives

$$\langle E_i^{(1)}(\mathbf{r}_1)E_i^{(2)}(\mathbf{r}_1)\rangle = -\frac{\langle E_j\rangle\langle E_k\rangle}{(4\pi\langle\epsilon\rangle)^3}\int_V d\omega_{12}\int_V d\omega_{23}\int_V d\omega_{34} \times \frac{x_{12,i}}{r_{12}^3}\frac{x_{13,i}}{r_{13}^3}\frac{x_{34,k}}{r_{34}^3}\frac{\partial^3\langle\epsilon'(\mathbf{r}_2)\epsilon'(\mathbf{r}_3)\epsilon'(\mathbf{r}_4)\rangle}{\partial x_{23,j}\partial x_{23,h}\partial x_{34,k}}. \quad (A2)$$

In spherical coordinates we have

$$\int_V d\omega_{12}\frac{x_{12,i}}{r_{12}^3}\frac{x_{13,i}}{r_{13}^3} = -\int_V \frac{d\omega_{12}}{r_{12}^2}\frac{x_{12,i}}{r_{12}}\frac{\partial}{\partial x_{12,i}}\left(\frac{1}{r_{13}}\right) = -\int_0^{2\pi} d\phi_{12}\int_0^{\pi} d\theta_{12}\sin\theta_{12}\int_0^{\infty} dr_{12} \times \frac{\partial x_{12,i}}{\partial r_{12}}\frac{\partial}{\partial x_{12,i}}\left(\frac{1}{r_{13}}\right) = \frac{4\pi}{r_{23}}. \quad (A3)$$

Therefore, Eq. (A2) becomes

$$\langle E_i^{(1)}(\mathbf{r}_1)E_i^{(2)}(\mathbf{r}_1)\rangle = -\frac{\langle E_j\rangle\langle E_k\rangle}{(4\pi)^2\langle\epsilon\rangle^3}\int_V d\omega_{23}\int_V d\omega_{34}\frac{1}{r_{23}}\frac{x_{34,k}}{r_{34}^3} \times \frac{\partial^3\langle\epsilon'(\mathbf{r}_2)\epsilon'(\mathbf{r}_3)\epsilon'(\mathbf{r}_4)\rangle}{\partial x_{23,j}\partial x_{23,h}\partial x_{34,k}}. \quad (A4)$$

For the three-point correlation function $f(\mathbf{r}_{32}, \mathbf{r}_{34})$, we impose boundary conditions such that as $r_{32} \to \infty$

$$\frac{\partial f}{\partial r_{32}} = o\left(\frac{1}{r_{32}}\right), \quad \frac{\partial f}{\partial \theta_{32}} = o(1), \quad \frac{\partial f}{\partial \phi_{32}} = o(1). \tag{A5}$$

Then, after integrating by parts and applying Gauss' theorem, we can see that

$$\begin{aligned}\langle E_i^{(1)}(\mathbf{r}_1)E_i^{(2)}(\mathbf{r}_1)\rangle &= -\frac{\langle E_j\rangle\langle E_k\rangle}{(4\pi)^2}\frac{\langle\epsilon'^3\rangle}{\langle\epsilon\rangle^3}\int_V d\omega_{32} \\ &\times \int_V d\omega_{34}\frac{x_{32,j}}{r_{32}^3}\frac{x_{34,h}}{r_{34}^3}\frac{\partial^2 f(\mathbf{r}_{32},\mathbf{r}_{34})}{\partial x_{32,h}\partial x_{34,k}} \\ &= -A_{jk}^{(3)}\frac{\langle\epsilon'^3\rangle}{\langle\epsilon\rangle^3}\langle E_j\rangle\langle E_k\rangle. \end{aligned} \tag{A6}$$

The evaluation of $\langle E_i^{(n)}E_i^{(m)}\rangle$ for any given n and m proceeds analogously. The result is

$$\langle E_i^{(n)}E_i^{(m)}\rangle = (-1)^{n+m}A_{jk}^{(n+m)}(\langle\epsilon'^{n+m}\rangle/\langle\epsilon\rangle^{n+m})\langle E_j\rangle\langle E_k\rangle. \tag{A7}$$

Lastly, we shall perform the proof of Eq. (3.10). Again we treat the simple case when $n = 1$ and $m = 2$. By definition,

$$\begin{aligned}&\langle\epsilon'(\mathbf{r}_2)E_i^{(1)}(\mathbf{r}_2)E_i^{(2)}(\mathbf{r}_2)\rangle \\ &= \frac{\langle E_j\rangle\langle E_k\rangle}{(4\pi\langle\epsilon\rangle)^3}\int_V d\omega_{21}\int_V d\omega_{23}\int_V d\omega_{34}\frac{x_{21,i}}{r_{21}^3}\frac{x_{23,i}}{r_{23}^3}\frac{x_{34,h}}{r_{34}^3} \\ &\times\frac{\partial^3\langle\epsilon'(\mathbf{r}_1)\epsilon'(\mathbf{r}_2)\epsilon'(\mathbf{r}_3)\epsilon'(\mathbf{r}_4)\rangle}{\partial x_{21,j}\partial x_{23,h}\partial x_{34,k}} \\ &= \frac{\langle E_j\rangle\langle E_k\rangle}{(4\pi\langle\epsilon\rangle)^3}\int_V d\omega_{12}\int_V d\omega_{23}\int_V d\omega_{34}\frac{x_{12,i}}{r_{12}^3}\frac{x_{23,i}}{r_{23}^3}\frac{x_{34,h}}{r_{34}^3} \\ &\times\frac{\partial^3\langle\epsilon'(\mathbf{r}_1)\epsilon'(\mathbf{r}_2)\epsilon'(\mathbf{r}_3)\epsilon'(\mathbf{r}_4)\rangle}{\partial x_{12,j}\partial x_{23,h}\partial x_{34,k}}. \end{aligned} \tag{A8}$$

Under appropriate boundary conditions like (A5), we obtain

$$\begin{aligned}&\int_V d\omega_{12}\frac{x_{12,i}}{r_{12}^3}\frac{\partial f(\mathbf{r}_{12},\mathbf{r}_{13},\mathbf{r}_{14})}{\partial x_{12,j}} \\ &= \int_V d\omega_{12}\frac{1}{r_{12}}\frac{\partial^2 f(\mathbf{r}_{12},\mathbf{r}_{13},\mathbf{r}_{14})}{\partial x_{12,i}\partial x_{12,j}} \\ &= \int_V d\omega_{12}\frac{x_{12,j}}{r_{12}^3}\frac{\partial f(\mathbf{r}_{12},\mathbf{r}_{13},\mathbf{r}_{14})}{\partial x_{12,i}}. \end{aligned} \tag{A9}$$

Introduction of Eq. (A9) into Eq. (A8) leads to

$$\langle\epsilon'(\mathbf{r}_2)E_i^{(1)}(\mathbf{r}_2)E_i^{(2)}(\mathbf{r}_2)\rangle = -A_{jk}^{(4)}(\langle\epsilon'^4\rangle/\langle\epsilon\rangle^3)\langle E_j\rangle\langle E_k\rangle. \tag{A10}$$

More generally,

$$\langle\epsilon' E_i^{(n)}E_i^{(m)}\rangle = (-1)^{n+m}A_{jk}^{(n+m+1)}(\langle\epsilon'^{n+m+1}\rangle/\langle\epsilon\rangle^{n+m})\langle E_j\rangle\langle E_k\rangle. \tag{A11}$$

APPENDIX B: EVALUATION OF $\langle\kappa' E_i^{(n)}\rangle$ AND $\langle\kappa' E_i^{(n)}E_i^{(m)}\rangle$

The proof of Eqs. (4.9) and (4.10) goes in a similar manner to that of Eqs. (3.8)–(3.10). Since the validity of Eq. (4.9) is almost self-evident, we shall restrict ourselves to the derivation of Eq. (4.10).

Let us estimate $\langle\kappa' E_i^{(1)}E_i^{(2)}\rangle$ in analogy to the arguments advanced in Appendix A. Using Eq. (3.2) and changing coordinates, we can write

$$\begin{aligned}&\langle\kappa'(\mathbf{r}_2)E_i^{(1)}(\mathbf{r}_2)E_i^{(2)}(\mathbf{r}_2)\rangle \\ &= \frac{\langle E_j\rangle\langle E_k\rangle}{(4\pi\langle\epsilon\rangle)^3}\int_V d\omega_{12}\int_V d\omega_{23}\int_V d\omega_{34}\frac{x_{12,i}}{r_{12}^3}\frac{x_{23,i}}{r_{23}^3}\frac{x_{34,h}}{r_{34}^3} \\ &\times\frac{\partial^3\langle\epsilon'(\mathbf{r}_1)\kappa'(\mathbf{r}_2)\epsilon'(\mathbf{r}_3)\epsilon'(\mathbf{r}_4)\rangle}{\partial x_{12,j}\partial x_{23,h}\partial x_{34,k}}. \end{aligned} \tag{B1}$$

As boundary conditions prescribed for $g_2(\mathbf{r}_{12}, \mathbf{r}_{13}, \mathbf{r}_{14})$, we assume that as $r_{12} \to \infty$

$$\frac{\partial g_2}{\partial r_{12}} = o\left(\frac{1}{r_{12}}\right), \quad \frac{\partial g_2}{\partial \theta_{12}} = o(1), \quad \frac{\partial g_2}{\partial \phi_{12}} = o(1), \tag{B2}$$

so that

$$\begin{aligned}&\int_V d\omega_{12}\frac{x_{12,i}}{r_{12}^3}\frac{\partial g_2(\mathbf{r}_{12},\mathbf{r}_{13},\mathbf{r}_{14})}{\partial x_{12,j}} \\ &= \int_V d\omega_{12}\frac{x_{12,j}}{r_{12}^3}\frac{\partial g_2(\mathbf{r}_{12},\mathbf{r}_{13},\mathbf{r}_{14})}{\partial x_{12,i}}. \end{aligned} \tag{B3}$$

Substitution of Eq. (B3) into Eq. (B1) yields

$$\langle\kappa' E_i^{(1)}E_i^{(2)}\rangle = -B_{jk}^{(4,2)}(\langle\kappa'\epsilon'^3\rangle/\langle\epsilon\rangle^3)\langle E_j\rangle\langle E_k\rangle, \tag{B4}$$

which is easily generalized to

$$\begin{aligned}\langle\kappa' E_i^{(n)}E_i^{(m)}\rangle &= (-1)^{n+m}B_{jk}^{(n+m+1,n+1)} \\ &\times(\langle\kappa'\epsilon'^{n+m}\rangle/\langle\epsilon\rangle^{n+m})\langle E_j\rangle\langle E_k\rangle. \end{aligned} \tag{B5}$$

APPENDIX C: PROOF OF $M_{ij}^{(3)} \leq 1$

To deduce $M_{ii}^{(3)} \leq 1$ we start from the inequality (3.16)

$$A_i^{(2)}\langle\epsilon'^2\rangle/\langle\epsilon\rangle^2 + A_i^{(3)}\langle\epsilon'^3\rangle/\langle\epsilon\rangle^3 \geq 0. \tag{C1}$$

Inserting Eqs. (5.4) and (5.5) into Eq. (C1) we find

$$-(\langle\epsilon'^3\rangle/\langle\epsilon'^2\rangle\langle\epsilon\rangle)M_{ii}^{(3)} \leq 1. \tag{C2}$$

Whatever the value of such a one-point moment as $\langle\epsilon\rangle$, $\langle\epsilon'^2\rangle$, or $\langle\epsilon'^3\rangle$ may be, this inequality must be always valid. On the other hand, it is evident that

$$-\langle\epsilon'^3\rangle/\langle\epsilon'^2\rangle\langle\epsilon\rangle = 1 - \langle\epsilon'^2\epsilon\rangle/\langle\epsilon'^2\rangle\langle\epsilon\rangle \leq 1. \tag{C3}$$

For a symmetric cell material composed of two phases with permittivities ϵ_1 and ϵ_2, the equality sign in Eq. (C3) holds actually in the limit as $\epsilon_1/\epsilon_2 \to 0$ and $v_1 \to 0$. It is thus concluded that

$$M_{ii}^{(3)} \leq 1. \tag{C4}$$

[1] M. Hori, J. Math. Phys. 14, 514 (1973).
[2] O. Wiener, Abh. Math.-Physik. Kl. Kgl. Sächs. Ges. 32, 509 (1912).
[3] Z. Hashin and S. Shtrikman, J. Appl. Phys. 33, 3125 (1962).
[4] Z. Hashin and S. Shtrikman, Phys. Rev. 130, 129 (1963).
[5] W. F. Brown Jr., Trans. Soc. Rheol. 9, 357 (1965).
[6] M. Beran, Nuovo Cimento 38, 771 (1965).
[7] M. Beran and J. Molyneux, Quart. Appl. Math. 24, 107 (1966).
[8] M. N. Miller, J. Math. Phys. 10, 1988 (1969).
[9] M. N. Miller, J. Math. Phys. 10, 2005 (1969); 12, 1057 (1971).
[10] M. J. Beran and N. R. Silnutzer, J. Compos. Mater. 5, 246 (1971).
[11] M. J. Beran, *Statistical Continuum Theories* (Interscience, New York, 1968), p. 240.
[12] W. F. Brown Jr., *Magnetostatic Principles in Ferromagnetism* (North-Holland, Amsterdam, 1962), p. 49.
[13] R. Moskowitz and E. Della Torre, IEEE Trans. Magn. 2, 739 (1966).

On the improper neglect of certain terms in random function theory

I. Lerche* and E. N. Parker†

Department of Physics, University of Chicago, Chicago, Illinois 60637
(Received 2 April 1973)

This paper presents some exact solutions of problems in random function theory for the purpose of testing the validity of an approximate method known variously in the many different fields of its application as first order smoothing theory, first order cumulant discard, quasilinear theory, or the adiabatic approximation. The hydromagnetic dynamo equations are used here, as particularly appropriate for such an investigation. The calculations show that in one case the exact and approximate solutions agree. In the other case the approximate solution is wrong. Hence, in the absence of a general criterion for validity, a result based on first order smoothing theory is a conjecture rather than a fact. This impacts strongly on much of the recent work on hydromagnetic dynamos.

I. INTRODUCTION

The formal theoretical basis for the hydromagnetic dynamo equations has come under close scrutiny recently[1] particularly in connection with the difficult question of whether statistically homogeneous isotropic turbulence in a conducting fluid of infinite extent leads to the growth of magnetic field. Of particular concern is the popular mathematical maneuver known variously as first order smoothing theory, first order cumulant discard, quasilinear theory, or the adiabatic approximation,[2-4] which is the basis for some of the more formal work on hydromagnetic dynamos, plasma turbulence, etc. First order smoothing theory conjectures that the terms which are nonlinear in the random functions do not differ much, or for very long, from their mean values, so that the difference can be neglected. The general validity of first order smoothing has not been established.

A number of authors have been concerned about the validity of the method, which neglects all mode–mode coupling (see review by Frisch[4] and references therein). A number of authors[5-8] have shown that mode–mode coupling is important in the approach to equilibrium in many cases in plasma physics, although there are also some situations in which the coupling can apparently be neglected with impunity.[9,10] Kraichnan[11] has established the invalidity of first order smoothing in special cases.[12] Herring[14] gives detailed comparisons of the quasilinear, the quasinormal, and Kraichnan's direct interaction approximation with numerical solutions of Boussinesq convection, and finds that only the direct interaction approximation agrees closely with the numerical solutions.

Our own interest in the problem originally arose in connection with the hydromagnetic dynamo and the origin of large-scale turbulent and ordered magnetic fields in astrophysical bodies. Consequently, the mathematical examples which we present are chosen from that domain.

Consider, then, a random velocity field $\delta v_i(\mathbf{r}, t)$, $\langle \delta v_i \rangle = 0$ interacting with a magnetic field $B_i(\mathbf{r}, t) + \delta B_i(\mathbf{r}, t)$, where $B_i(\mathbf{r}, t)$ is the ensemble average field, presumably nonvanishing almost everywhere, and δB_i is the random component, $\langle \delta B_i \rangle = 0$. The hydromagnetic equation for infinite conductivity $\delta \mathbf{B}/\partial t = \nabla \times (\mathbf{v} \times \mathbf{B})$ reduces to the equation

$$\frac{\partial B_i}{\partial t} = \frac{\partial}{\partial x_i} (\langle \delta v_i \delta B_i \rangle - \langle \delta v_i \delta B_i \rangle)$$

for the mean field B_i, and this equation is statistically *exact*. Subtracting this from the hydromagnetic equation yields the *exact* equation

$$\frac{\partial \delta B_i}{\partial t} = \frac{\partial}{\partial x_j} (\delta v_i B_j - \delta v_j B_i) + \frac{\partial}{\partial x_j} (\delta v_i \delta B_j - \langle \delta v_i \delta B_j \rangle) - \frac{\partial}{\partial x_j} (\delta v_j \delta B_i - \langle \delta v_j \delta B_i \rangle)$$

for the random component of the field. First order smoothing theory decrees that $\delta v_i \delta B_j$ never differs by much, or not for long, from $\langle \delta v_i \delta B_j \rangle$ so that the terms which are nonlinear in the random quantities can be neglected. Then we have the inexact equation

$$\frac{\partial \delta B_i}{\partial t} \cong \frac{\partial}{\partial x_j} (\delta v_i B_j - \delta v_j B_i),$$

which is then solved simultaneously with the exact equation for the average field.

We might expect the approximation to be valid when the correlation length and time of δv_i are both small. Whether it is valid under less restricted conditions has yet to be shown. Unfortunately, the method is often applied[15] without a general proof that it is accurate.

In a recent paper[16] we used first order smoothing theory to work out the small-scale turbulent components of the "azimuthal" field $B_y(x, t)$ and the vector potential $A(x, t)$ for the poloidal field $\partial A/\partial x$ in the z-direction as a consequence of random fluctuations in the large-scale shear G and in the cyclonic component Γ of the small-scale convection. The calculation was motivated by the galactic dynamo[17-20] in which one observes the mean of the total field $B_i + \delta B_i$ along a line of sight, and one then needs to know the magnitude of δB_i.

As was noted in that paper, an exact solution of the dynamo equations is possible, affording the opportunity to determine directly whether the results of first order smoothing theory are correct. This paper presents the exact solutions, and points out the success and failure, of first order smoothing.

In their simplest form (strong shear, rectangular geometry, and large dynamo number) the dynamo equations, describing the generation of field by shear and cyclonic turbulence, are[13,17]

$$\frac{\partial B_y}{\partial t} = G \frac{\partial A}{\partial x}, \tag{1}$$

$$\frac{\partial A}{\partial t} = \Gamma B_y, \tag{2}$$

where $G \equiv dV_y/dz$ represent the large-scale shear and the Γ represents the cyclonic velocity component. In this example the vector potential $A(x, t)$ is in the y-direction and the z component of field (sheared by G) is $\partial A/\partial x$.

We shall consider the irregularities in the field produced by a random variation of Γ. The dynamo equations give the large-scale (ensemble average) fields B_y and A in terms of the ensemble average shear and cyclonic rates G and Γ respectively. So imagine, then, an ensemble of systems in which the ensemble average cyclone strength Γ and/or shear G are functions of x and t. This ensemble constitutes our "microcanonical ensemble." We shall suppose that Γ and G (the microcanonical averages) are random functions of x and t. Then consider an ensemble of "microcanonical ensembles," essentially a "grand canonical ensemble." The dynamo equations (1) and (2) apply to each microcanonical ensemble. Consider their solution, using the grand ensemble for defining constant, or slowly varying, mean values, etc. Equations (1) and (2) can be solved *exactly* when either Γ or G is a constant and the other is a function of either x or t (sec. II and III).

II. Γ OR G A RANDOM FUNCTION OF TIME

A. General considerations

Consider the two cases

$$\Gamma = \Gamma_0, \quad G(t) = G_0[1 + \epsilon\delta G(t)], \tag{3}$$

$$\Gamma = \Gamma_0[1 + \epsilon\delta\Gamma(t)], \quad G = G_0, \tag{4}$$

where Γ_0 and G_0 are constants and $\delta\Gamma(t)$ and $\delta G(t)$ have zero mean values, $\langle\delta\Gamma\rangle = \langle\delta G\rangle = 0$. Then the two dynamo Eqs. (1) and (2) can be written

$$\frac{\partial^2 A}{\partial t^2} - \Gamma_0 G_0 (1 + \epsilon\delta G)\frac{\partial A}{\partial x} = 0 \tag{5}$$

and

$$\frac{\partial^2 B_y}{\partial t^2} - \Gamma_0 G_0 (1 + \epsilon\delta\Gamma)\frac{\partial B_y}{\partial x} = 0 \tag{6}$$

for (3) and (4) respectively. The equations are identical in form, so that it is sufficient to consider (6) alone. Henceforth we shall drop the subscripts zero and understand that Γ and G denote the constant parts of the cyclonic strength and shear, respectively. Let $h = t/T$, where T is the correlation time for $\delta\Gamma(t)$. Then let

$$B_y = B(h)\exp ikx. \tag{7}$$

Equation (6) reduces to

$$\frac{d^2 B}{dh^2} - ik\Gamma G T^2[1 + \epsilon\delta\Gamma(h)]B = 0. \tag{8}$$

This equation is second order and may be written as two first order equations if we define

$$Y \equiv B, \quad Z \equiv \frac{dB}{dh}. \tag{9}$$

Then

$$\frac{dZ}{dh} = i\Gamma G T^2(1 + \epsilon\delta\Gamma)Y, \tag{10}$$

$$\frac{dY}{dh} = Z \tag{11}$$

Consider the probability $P(h, \delta\Gamma, Y, Z)$ for finding the combination of values $(\delta\Gamma, Y, Z)$. Assume that the probability distribution $\int dY \int dZ P$ over $\delta\Gamma$ alone is gaussian. Then $P(h, \delta\Gamma, Y, Z)$ satisfies the spatially homogeneous probability equation[21]

$$\frac{\partial P}{\partial h} = \frac{\partial}{\partial\delta\Gamma}\delta\Gamma P + \frac{\partial^2 P}{\partial\delta\Gamma^2} - \frac{\partial}{\partial Y}\left(\frac{dY}{dh}P\right) - \frac{\partial}{\partial Z}\left(\frac{dZ}{dh}P\right), \tag{12}$$

with dY/dh and dZ/dh given by (10) and (11). The first two terms on the right hand side represent the assumption that the probability distribution over $\delta\Gamma$ is the Gaussian $\exp(-\frac{1}{2}\delta\Gamma^2)$. The initial values of B and dB/dh are sufficient to determine the solution of (8). Denote them by $Y(0)$ and $Z(0)$, respectively, so that when $h = 0 = t$ the probability distribution is

$$P_0 = (2\pi)^{-1/2}\exp(-\tfrac{1}{2}\delta\Gamma^2)\,\delta[Y - Y(0)]\delta[Z - Z(0)]. \tag{13}$$

To proceed with the solution of (12), define the quantities

$$R(h, \delta\Gamma) \equiv \int dY \int dZ\, YP,$$

$$S(h, \delta\Gamma) \equiv \int dY \int dZ\, ZP,$$

so that

$$\langle Y\rangle = \langle B\rangle = \int_{-\infty}^{+\infty} d\delta\Gamma\, R(h, \delta\Gamma),$$

$$\langle Z\rangle = \left\langle\frac{dB}{dh}\right\rangle = \int_{-\infty}^{+\infty} d\delta\Gamma\, S(h, \delta\Gamma).$$

Then (12) yields

$$\frac{\partial R}{\partial h} = \frac{\partial}{\partial\delta\Gamma}(R\delta\Gamma) + \frac{\partial^2 R}{\partial\delta\Gamma^2} + S \tag{14}$$

and

$$\frac{\partial S}{\partial h} = \frac{\partial}{\partial\delta\Gamma}(S\delta\Gamma) + \frac{\partial^2 S}{\partial\delta\Gamma^2} + ik\Gamma G(1 + \epsilon\delta\Gamma)R, \tag{15}$$

and the initial conditions

$$R(0, \delta\Gamma) = Y(0)(2\pi)^{-1/2}\exp(-\tfrac{1}{2}\delta\Gamma^2),$$

$$S(0, \delta\Gamma) = S(0)(2\pi)^{-1/2}\exp(-\tfrac{1}{2}\delta\Gamma^2).$$

The coefficients in (14) and (15) are independent of time h, so that the solutions have an exponential time dependence $\exp i\nu h$. Note that (14) and (15) are homogeneous equations in R and S. Hence they have a solution if, and only if, some dispersion relation is satisfied. Our purpose is to obtain that dispersion relation.

It is convenient to expand R and S in the normal modes ψ_n of the homogeneous equation

$$\frac{d^2\psi_n}{d\delta\Gamma^2} + \frac{d}{d\delta\Gamma}\delta\Gamma\psi_n + n\psi_n = 0,$$

which are

$$\psi_n = \exp(-\tfrac{1}{2}\delta\Gamma^2)H_n(2^{-1/2}\delta\Gamma), \tag{16}$$

where H_n is the nth Hermite polynomial. Write $\exp i\nu h = \exp i\omega t$ where $\nu \equiv \omega T$. Then consider the solutions

$$R = \exp i\nu h \sum_{n=0}^{\infty} C_n\psi_n(\delta\Gamma),$$

$$S = \exp i\nu h \sum_{n=0}^{\infty} D_n\psi_n(\delta\Gamma),$$

and substitute into (14) and (15). Noting that

$$2^{1/2}\delta\Gamma\psi_n(\delta\Gamma) = \psi_{n+1}(\delta\Gamma) + 2n\psi_{n-1}(\delta\Gamma) \tag{17}$$

and equating coefficients of the ψ_n gives the two equations

$$(i\nu + n)C_n - D_n = 0,$$

$$-ik\Gamma GT^2[C_n + 2^{-1/2}\epsilon C_{n-1} + 2^{1/2}\epsilon(n+1)C_{n+1}] + (i\nu + n)D_n = 0.$$

Then write the dimensionless wavenumber $\alpha \equiv k\Gamma GT^2$ and eliminate D_n, obtaining

$$2^{-1/2}\alpha\epsilon C_{n-1} + [\alpha + i(i\nu + n)^2]C_n + 2^{1/2}\alpha\epsilon(n+1)C_{n+1} = 0. \tag{18}$$

B. Solution by determinant

The determinant of the coefficients of (18) gives the dispersion relation. The determinant is infinite, and divergent. However, expanding the determinant about the upper left hand corner gives a series that appears to be asymptotically convergent for small $\epsilon\alpha$. The 2×2 determinant gives the first correction

$$[\alpha + i(i\nu)^2][\alpha + i(i\nu + 1)^2] = \alpha^2\epsilon^2. \tag{19}$$

The original roots, arising from the dynamo equations when $\epsilon = 0$, are

$$\alpha + i(i\nu)^2 = 0$$

or

$$i\nu = \pm\, \alpha^{1/2} \exp i\pi/4$$

(one growing and one decaying when k is real). There is now an additional pair of roots

$$i\nu = -1 \pm \alpha^{1/2} \exp i\pi/4.$$

Had we evaluated the 3×3 determinant, there would appear still another pair

$$i\nu = -2 \pm \alpha^{1/2} \exp i\pi/4$$

etc. The extra roots all converge to the original roots in the limit as $T \to \infty$. The real part of $i\nu$ (the growth rate) is smaller for these additional roots than in the original pair, and so we will not consider them further in the present problem.

Consider the question of the convergence of (19) giving the correction $O(\epsilon^2)$ to the original dispersion relation $\alpha + i(i\nu)^2$. Work out the 3×3 determinant, keeping terms $O(\epsilon^4)$. Use (19) to eliminate $i\nu$ from the correction terms. The result can be written

$$\alpha + i(i\nu)^2 = \alpha^2\epsilon^2/[\alpha + i(i\nu + 1)^2] - 2\alpha^4\epsilon^4/[\alpha + i(i\nu + 1)^2][\alpha + i(i\nu + 2)^2] + O(\alpha^6\epsilon^6).$$

If we now work out the 4×4 determinant, and then the 5×5 determinant, keeping terms $O(\epsilon^4)$ and using (19) to eliminate $i\nu$ in the correction terms, we obtain no further contributions to the terms $O(\epsilon^2)$ and $O(\epsilon^4)$ etc. Hence the expansion of the determinant generates a series which appears to be asymptotically convergent for small α and/or ϵ. We shall check this assertion later.

For present purposes it is sufficient to consider only the first correction terms $O(\epsilon^2)$. By writing $\nu = \omega T$, so that the time dependence is $\exp(i\omega t)$, it is readily shown that the growing mode is

$$i\omega T = (\alpha/2)^{1/2}\{[1 - \tfrac{1}{2}\alpha\epsilon^2(1 + 2^{3/2}\alpha^{1/2} + 4\alpha)^{-1}] + i[1 + \tfrac{1}{2}\alpha\epsilon^2(1 + 2^{3/2}\alpha^{1/2})(1 + 2^{3/2}\alpha^{1/2} + 4\alpha)^{-1}]\}. \tag{20}$$

Then if $\alpha \ll 1$,

$$i\omega T \simeq (\alpha/2)^{1/2}[(1 - \tfrac{1}{2}\alpha\epsilon^2) + i(1 + \tfrac{1}{2}\alpha\epsilon^2)], \tag{21}$$

and if $\alpha \gg 1$,

$$i\omega T \simeq (\alpha/2)^{1/2}[(1 - \tfrac{1}{8}\epsilon^2) + i(1 + 2^{-3/2}\alpha^{1/2}\epsilon^2)]. \tag{22}$$

Thus when Γ is a random function of t (or G a random function of t) the real part of $i\omega T$ (the growth rate) is reduced and the imaginary part (the phase velocity) is increased.

C. Solution by differential equation

A more proper calculation of the dispersion relation, which does not rely on throwing away the divergent terms in an infinite determinant, is to write the coefficient C_n as

$$C_n = \int_0^{2\pi} d\beta \exp(in\beta)\, c(\beta)$$

Then $c(\beta)$ is related to C_n by the Fourier series

$$c(\beta) = \frac{1}{2\pi}\sum_{n=0}^{\infty} C_n \exp(-in\beta),$$

and is obviously a periodic function of β with period 2π. Then (18) can be written

$$\frac{d^2c}{d\beta^2} + 2(\nu - \alpha\Delta \exp i\beta)\frac{dc}{d\beta} + \{\nu^2 + i\alpha[1 + \Delta \exp(-i\beta)]\}\, c = 0, \tag{23}$$

where we have written $\Delta \equiv 2^{-1/2}\epsilon$.

Now, if $\epsilon = 0$, the solution of (18) is elementary; if $C_n \neq 0$, then

$$\alpha + i(i\nu + n)^2 = 0,$$

which we recognize as the sequence of roots appearing in the $n \times n$ determinant (19). The solution of the dynamo equations (1) and (2) for $\epsilon = 0$ gives

$$i\nu = \alpha^{1/2} \exp i\pi/4,$$

corresponding to $C_0 = O(1)$, say. For small but non-vanishing ϵ we expect that $C_0 = O(1)$ and hence that

$$C_0 = \int_0^{2\pi} d\beta\, c(\beta) = O(1).$$

With this in mind, consider the solution of (23). Write $dc/d\beta = c\Phi$ to reduce (23) to the Riccati equation. Then expand Φ as

$$\Phi(\beta) = \sum_{n=0}^{\infty} \Delta^n\Phi_n(\beta)$$

and equate the coefficients of like powers of Δ,

$$\Phi_0' + \Phi_0^2 + 2\nu\Phi_0 + \nu^2 + i\alpha = 0,$$

$$\Phi_1' + 2\Phi_1(\Phi_0 + \nu) + i2^{-1/2}\alpha \exp(-i\beta) - 2^{1/2}\alpha\Phi_0 \exp i\beta = 0,$$

$$\Phi_2' + 2\Phi_2(\Phi_0 + \nu) + \Phi_1^2 - 2^{1/2}\alpha\Phi_1 \exp i\beta = 0,$$

etc. Before setting out to solve this system of equations, we must consider in a little more detail the nature of the solutions. Since $c(\beta)$ is periodic and

$$c(\beta) = \exp \sum \Delta^n \int^{\beta} du\Phi_n(u),$$

it is evident that each $\int^{\beta} du\Phi_n(u)$ must consist solely of terms linear in β and perhaps periodic terms, such as $\exp(\pm i\beta)$. Hence each Φ_n must consist only of a constant term plus periodic terms.

The equation for Φ_0 gives no periodic terms $\exp\pm i\beta$ so Φ_0 must be only a constant, yielding $\Phi_0' = 0$ and

$$\Phi_0 = -\nu \pm \alpha^{1/2} \exp(-i\pi/4).$$

There will be constant terms in Φ_1, Φ_2, etc., too. The terms Φ_1, Φ_2, etc. will contain periodic terms, so that their sum is schematically

$$\Phi = C + \epsilon D \exp(\pm i\beta) + O(\epsilon^2)$$

and

$$c(\beta) \cong \exp[C\beta \mp i\epsilon D \exp(\pm i\beta)].$$

But $c(\beta)$ is periodic in β. Hence C must be equal to mi, where m is an integer. Hence

$$C_0 = \int_0^{2\pi} d\beta \exp[im\beta \mp i\epsilon D \exp(\pm i\beta)].$$

But if $m = \pm 1, \pm 2$, etc. this integral for C_0 is only $O(\epsilon)$, whereas we pointed out above that $C_0 = O(1)$. Therefore, m cannot be $\pm 1, \pm 2$, etc. but must be zero, i.e., the sume of the constant terms in $\sum \Delta^n\Phi_n$ must vanish. This restriction provides the dispersion relation.

It is readily shown that

$$\Phi_1 = \frac{2^{1/2}\alpha\Phi_0 \exp i\beta}{[i + 2(\Phi_0 + \nu)]} - \frac{i\alpha}{2^{1/2}} \frac{\exp(-i\beta)}{[i - 2(\Phi_0 + \nu)]}$$

so that Φ_1 is wholly periodic. For Φ_2 we have

$$\Phi_2 = -\frac{i\alpha^2(i + 2\nu + \Phi_0)}{2(\Phi_0 + \nu)[1 + 4(\Phi_0 + \nu)^2]} + \text{periodic terms}.$$

The sum of the constant terms through $O(\Delta^2)$ must vanish. Solving for $i\nu(\equiv i\omega T)$ yields

$$i\nu = \alpha^{1/2} \exp(i\pi/4) \left(1 + \frac{\alpha\epsilon^2[i + 2\alpha^{1/2} \exp(-i\pi/4)]}{2(1 - 4i\alpha)}\right),$$

which is, of course, identical with (20) and (21).

III. Γ OR G A RANDOM FUNCTION OF SPACE

A. General considerations

Consider the two cases

$$\Gamma = \Gamma_0, \quad G(x) = G_0[1 + \epsilon\delta G(x)], \tag{24}$$

$$\Gamma(x) = \Gamma_0[1 + \epsilon\delta\Gamma(x)], \quad G = G_0, \tag{25}$$

where Γ_0 and G_0 are constants and $\delta\Gamma$ and δG have zero mean. Then the dynamo equations (1) and (2) can be written

$$\frac{\partial^2 B_y}{\partial t^2} - \Gamma_0 G_0[1 + \epsilon\delta G(x)] \frac{\partial B_y}{\partial x} = 0, \tag{26}$$

and

$$\frac{\partial^2 A}{\partial t^2} - \Gamma_0 G_0[1 + \epsilon\delta\Gamma(x)] \frac{\partial A}{\partial x} = 0 \tag{27}$$

for (24) and (25) respectively. The equations are identical in form, so the two cases are equivalent. It is sufficient to consider only (27). Again we drop the subscripts, zero on Γ_0 and G_0.

Equation (27) admits of exact solutions[19] by elementary methods, but it is more useful to obtain the dispersion relation using the techniques of the previous section.

Let

$$A(x, t) = \exp(i\omega t)F(x).$$

Then with $y = x/L$, (27) becomes

$$\frac{dF}{dy} = -\frac{\omega^2 LF}{\Gamma G(1 + \epsilon\delta\Gamma)} \tag{28}$$

Consider the joint probability distribution $P(y, F, \delta F)$, satisfying[21]

$$\frac{\partial P}{\partial y} = \frac{\partial}{\partial \delta\Gamma} \delta\Gamma P + \frac{\partial^2 P}{\partial \delta\Gamma^2} - \frac{\partial}{\partial F}\left(\frac{dF}{dy} P\right).$$

Define the energy distribution

$$U(\delta\Gamma) = \int dF\, FP$$

so that

$$\langle F\rangle = \int d\delta\Gamma\, U(\delta\Gamma).$$

Then

$$\frac{\partial U}{\partial y} = \frac{\partial}{\partial \delta\Gamma} \delta\Gamma U + \frac{\partial^2 U}{\partial \delta\Gamma^2} - \frac{\omega^2 LU}{\Gamma G(1 + \epsilon\delta\Gamma)} \tag{29}$$

with dF/dy given by (28). Again expand U in terms of $\psi_n(\delta\Gamma)$, defined in (16),

$$U(y, \delta\Gamma) = \exp(iqy) \sum_{n=0}^{\infty} E_n\psi_n(\delta\Gamma).$$

Substitute into (29) and note the identity (17) for $\delta\Gamma\psi_n(\delta\Gamma)$. Equating coefficients of ψ_n leads to the set of equations

$$2^{-1/2}\epsilon(iq + n - 1)E_{n-1} + \left(iq + \frac{\omega^2 L}{\Gamma G} + n\right)E_n + 2^{1/2}\epsilon(iq + n + 1)(n + 1)E_{n+1} = 0.$$

B. Solution by determinant

The determinant of the coefficients gives the desired dispersion relation. The determinant is infinite and divergent, but for $\epsilon \ll 1$ it is asymptotically convergent, as in the previous case. The lowest order term is the element in the upper left-hand corner. Equating this element to zero gives the dispersion relation.

$$\omega^2 + ik\Gamma G = 0$$

for $q \equiv kL$ and $\epsilon = 0$. The lowest 2×2 determinant gives the correction $O(\epsilon^2)$. The lowest 3×3 determinant gives the correction to $O(\epsilon^4)$. The 4×4 determinant gives no further correction to the term $O(\epsilon^4)$, etc. Thus

$$\omega^2 = -ik\Gamma G\{1 - \epsilon^2(1 + ikL) - 2\epsilon^4(1 + ikL)^2 + O(\epsilon^6)\}.$$

If $k\Gamma G$ is real and positive, the growing wave is

$$i\omega \cong (\tfrac{1}{2}k\Gamma G)^{1/2}\{[1 - \tfrac{1}{2}\epsilon^2(1 - kL)] + i[1 - \tfrac{1}{2}\epsilon^2(1 + kL)]\}, \tag{30}$$

when Γ is a random function of x (or G a random function of x). Both the real and imaginary parts are reduced bv the random component ϵ.

C. Solution by differential equation

A more proper calculation of the dispersion relation, which does not involve direct questions of convergence of the determinant, is to write the coefficient E_n as

$$E_n = \int_0^{2\pi} d\beta \exp(in\beta)e(\beta)$$

and proceed as in the previous section, IIC. There results the differential equation

$$\frac{d^2e}{d\beta^2} + \left(q - \frac{i}{2^{1/2}\epsilon}\exp(-i\beta) - \frac{i}{2}\exp(-i2\beta)\right)\frac{de}{d\beta} - \left(\frac{iq + \omega^2 L/\Gamma G}{2^{1/2}\epsilon}\exp(-i\beta) + \tfrac{1}{2}iq\exp(-i2\beta)\right)e = 0.$$

Again write

$$\frac{de}{d\beta} = e\Psi,$$

giving a Riccati equation. Expand Ψ in powers of ϵ,

$$\Psi(\beta) = \sum_{n=0}^{\infty} \epsilon^n \Psi_n(\beta)$$

yielding

$$\Psi_0 = \frac{i\omega^2 L}{\Gamma G} - q,$$

$$\Psi_1 = -i2^{1/2}\Psi_0(\Psi_0 + q)\exp i\beta - 2^{-1/2}(\Psi_0 + q)\exp(-i\beta),$$

$$\Psi_2 = (\Psi_0 + q)[1 + i(q + 3\Psi_0)] + \ldots,$$

where the three dots indicate the terms of the form $\exp(\pm i2\beta)$, which are periodic in β. For the reasons mentioned in II C the part of Ψ which is independent of β must vanish, yielding the dispersion relation

$$\frac{\omega^2 L}{\Gamma G} + iq = -\epsilon^2\frac{\omega^2 L}{\Gamma G}\left(1 - 2iq - \frac{3\omega^2 L}{\Gamma G}\right) + Q(\epsilon^4) \cong i\epsilon^2 q(1 + iq) + O(\epsilon^4), \quad (31)$$

which reduces precisely to (30) for the growing mode.

IV. VALIDITY OF FIRST ORDER SMOOTHING?

Now first order smoothing [Ref. 16, Eq. (13)] gives

$$i\omega T = \alpha^{1/2}\exp(i\pi/4) \times \{1 - \tfrac{1}{2}\epsilon^2\alpha^{1/2}\beta(1 + \alpha^{1/2}\beta) \times [1 + \alpha^{1/2}\beta + \alpha^{1/2}\exp(i\pi/4)]^{-1} \times [1 + \alpha^{1/2}\beta - \alpha^{1/2}\exp(i\pi/4)]^{-1}\}, \quad (32)$$

where $\alpha^{1/2}$ is the dynamo frequency $(k\Gamma G)^{1/2}$ multiplied by T and

$$\beta \equiv [(1 + ikL)/kL]^{1/2}$$

for the case that G is constant and $\delta\Gamma$ is a function of both x and t, and the correlation function is

$$\langle\delta\Gamma(x', t')\delta\Gamma(x'', t'')\rangle = \exp[-|x' - x''|/L - (t' - t'')/T]$$

as would result from a Gaussian distribution of $\delta\Gamma$.

If $\delta\Gamma$ is a function only of x, then let $T \to \infty$ in (32) to remove the time dependence.[22] The result can be written

$$i\omega = (k\Gamma G)^{1/2}\exp(i\pi/4)[1 - \tfrac{1}{2}\epsilon^2(1 + ikL)] = (\tfrac{1}{2}k\Gamma G)^{1/2}\{[1 - \tfrac{1}{2}\epsilon^2(1 - kL)] + i[1 - \tfrac{1}{2}\epsilon^2(1 + kL)]\}.$$

This is identical with (30), showing that for all values of L first order smoothing gives a result in agreement to the order considered with the exact calculation.

If $\delta\Gamma$ is a function only of t, then let $L \to \infty$ to remove the space dependence from the general result (32). Then $\beta = \exp(i\pi/4)$ and (32) reduces to

$$i\omega T = \alpha^{1/2}\exp(i\pi/4) \times \{1 - \tfrac{1}{2}\epsilon^2\alpha^{1/2}\exp(i\pi/4) \times [1 + \alpha^{1/2}\exp(i\pi/4)][1 + 2\alpha^{1/2}\exp(i\pi/4)]^{-1}\} = (\tfrac{1}{2}\alpha)^{1/2}\{[1 - \tfrac{1}{2}\alpha\epsilon^2(1 + 2^{3/2}\alpha^{1/2} + 4\alpha)^{-1}] + i[1 - (\tfrac{1}{2}\alpha)^{1/2}\epsilon^2(1 + 3(\tfrac{1}{2}\alpha)^{1/2} + 2\alpha) \times (1 + 2^{3/2}\alpha^{1/2} + 4\alpha)^{-1}]\}. \quad (33)$$

The real part agrees exactly with (20). But the correction to the imaginary part differs in both sign and magnitude. We expect that first order smoothing theory should be best when the correlation T is small, i.e., $\alpha \ll 1$. In this limit

$$i\omega T \cong (\tfrac{1}{2}\alpha)^{1/2}(1 - \tfrac{1}{2}\alpha\epsilon^2 + i\{1 + \epsilon^2[-(\alpha/2)^{1/2} + \alpha/2]\}).$$

By comparing this with its counterpart (22) from the exact calculation it is apparent that the disagreement between (20) and (32) arises from the intrusion of an additional term $-\epsilon^2(\alpha/2)^{1/2}$ in the correction to the imaginary part computed from first order smoothing.

In the limit that $T \to 0$—the "short sudden" approximation[13,23] -we have

$$i\omega = (\tfrac{1}{2}k\Gamma G)^{1/2}(1 + i),$$

which is identical with the exact solution. But for any finite T first order smoothing disagrees in the lowest order terms in T with the exact solution.

Altogether, then, comparison of first order smoothing theory with the exact solutions using the methods of Uhlenbeck and Ornstein[21] gives agreement in one case but not in the other. We note, for what it is worth, that agreement arises when the differential equation [in this case (27)] is first order in the variable on which the random function depends, and disagreement when the differential equation [in this case (9)] is second order in the variable on which the random function depends. Evidently the cause of the disagreement is the nonlinear term $\delta\Gamma\delta B - \langle\delta\Gamma\delta B\rangle$ neglected in first order smoothing. The equation for the large-scale field is correct in first order smoothing provided only that the irreducible part of the triple correlation is zero. Evidently, in the one case, the triple correlation is not zero, and in the other is consistent with zero.

As far as we have been able to ascertain there has not yet been a systematic investigation of the conditions under which first order smoothing is an accurate approximation for treating random fluctuations. There do exist particular investigations (e.g., Kraichnan,[11] Lerche,[1] and this paper) which demonstrate its invalidity under conditions where it might have been believed, *a priori*, to be valid. There are also particular cases (e.g., this paper) that demonstrate its validity.

Thus, after completing a calculation in the first order smoothing approximation, one does not know whether the answer is close to the truth. The result *is* valid in the limit of short correlation length and time–the "short-sudden" approximation. But to show whether it is valid for finite correlation length and time, it is necessary to turn to the more sophisticated and more correct direct interaction approximation, developed by Kraichnan.

A number of fields of physical inquiry are affected. In particular, the work of Steenbeck, Krause, and Radler[24] (see also Moffatt,[25] Weiss,[26] and Krause and Roberts,[15] and the comments by Lerche,[27,28] and references therein) on hydromagnetic dynamos is based on the first order smoothing approximation. Their results can be judged valid only when they can be verified by accurate methods, such as the direct interaction approximation. Their results are, of course, valid in the short-sudden limit, which was first employed by us[13,23] in demonstrating the dynamo principle, and their results reduce precisely to our prior answers. Curiously enough Weiss[26] in his excellent review of the development of hydromagnetic dynamo theory refers to our original *formal* calculations, in the mathematically *valid* short-sudden limit, as "heuristic", evidently prefering the more formal recent work based on the first order smoothing *conjecture*. Fortunately we know that the more recent work is valid in the short sudden limit, where it reduces to the earlier results of Parker.[13]

Clearly a significant advance would be made in kinematic dynamo theory if the work of Steenback, Krause, and Radler could be extended (in a rigorous mathematical manner, of course) beyond the "short-sudden" regime. It is our opinion that the only valid mathematical tool available for the job is Kraichnan's direct interaction approximation.

*This portion of the work was supported in part by the United States Air Force, Air Force Systems Command under contract F-19628-72-C-0056.

†This portion of the work was supported in part by the National Aeronautics and Space Administration under NASA Grant NGL 14-001-001.

[1]I. Lerche, J. Math. Phys. **14**, 1381 (1973).
[2]A. M. Yaglom, *An Introduction to the Theory of Stationary Random Functions* (Prentice-Hall, Englewood Cliffs, N.J., 1962).
[3]W. C. Meecham, J. Geophys. Res. **69**, 3175 (1964).
[4]U. Frisch, "Wave Propagation in Random Media", in *Probabilistic Methods in Applied Mathematics*, Vol. 1, edited by A. T. Bharucha-Reid (Academic, New York, 1968).
[5]I. B. Bernstein and F. Engelmann, Phys. Fluids **9**, 937 (1966).
[6]A. Dolinsky and R. Goldman, Phys. Fluids **20**, 1251 (1967).
[7]B. B. Kadomtsev, *Plasma Turbulence* (Academic, New York, 1965).
[8]K. Y. Fu, Plasma Phys. **15**, 57 (1973).
[9]C. F. Kennel and F. Engelmann, Phys. Fluids **9**, 2377 (1966).
[10]C. F. Kennel and H. E. Petschek, J. Geophys. Res. **71**, 1 (1966).
[11]R. H. Kraichnan, J. Math. Phys. **2**, 124 (1961).
[12]The reader is referred to a paper by Lerche[1] for a comparison of the results of first order smoothing with the results of Kraichnan's equations, which latter represent, in a statistically *exact* manner, an ensemble of physically possible dynamical systems. Exact agreement obtains only under Parker's[13] "short-sudden" conditions.
[13]E. N. Parker, Astrophys. J. **122**, 293 (1955).
[14]J. R. Herring, Phys. Fluids **12**, 39 (1969).
[15]F. Krause and P. H. Roberts, Astrophys. J. **181**, 977 (1973).
[16]I. Lerche and E. N. Parker, Astrophys. J. **168**, 231 (1971).
[17]E. N. Parker, Astrophys. J. **162**, 665 (1970).
[18]E. N. Parker, Astrophys. J. **163**, 295 (1971).
[19]E. N. Parker, Astrophys. J. **166**, 295 (1971).
[20]E. N. Parker, Astrophys. J. **168**, 239 (1971).
[21]G. E. Uhlenbeck and L. S. Ornstein, Phys. Rev. **36**, 823 (1930).
[22]The validity of first order smoothing may be questionable when T (or L) is comparable to, or larger than, the characteristic time or scale of the ordered field. But passing to the limit $T \to \infty$ (or $L \to \infty$) in the present case gives the same result as is obtained if the calculation is repeated from the beginning with $\delta\Gamma$ independent of t (or x). Thus the correct dispersion relation is obtained, evidently, if $T \to \infty$ (or $L \to \infty$) in the framework of first order smoothing.
[23]E. N. Parker, Astrophys. J. **157**, 1119 (1969).
[24]M. Steenbeck, F. Krause, and K. H. Radler, Z. Naturforsch. **21a**, 369 (1966).
[25]K. Moffat, J. Fluid Mech. **41**, 435 (1970).
[26]N. O. Weiss, Quart J. Roy. Astron. Soc. **12**, 432 (1971).
[27]I. Lerche, Astrophys. J. **166**, 627 (1971).
[28]I. Lerche, Astrophys. J. **181**, 993 (1973).

Relation between the boson calculus and Zhelobenko's method

M. A. Lohe and C. A. Hurst

Department of Mathematical Physics, University of Adelaide, Adelaide, S. Australia 5001
(Received 30 May 1973)

A comparison is made between the method of constructing finite-dimensional representations of the classical groups using the boson calculus in its standard or modified form and the method of Zhelobenko, which uses polynomials over a homogeneous space defined by a certain triangular subgroup. It is shown that the two methods can be directly related, so that one construction can be mapped into the other.

1. INTRODUCTION

The method of the boson calculus has been extensively developed for the construction of finite-dimensional irreducible tensor representations of the unitary groups, and recently it was extended,[1] by introducing modified bosons, to do the same for the orthogonal and symplectic groups. As there is already in existence a fully developed formalism for constructing all these representations, together with the spinor representations of the orthogonal groups due to Zhelobenko,[2] it is of interest to establish the connection between the two. Some discussion of this relationship was given by Zhelobenko in the paper referred to, if the boson calculus is interpreted as the construction of irreducible tensor representations, but a detailed correspondence showing exactly how to relate the polynomial bases which occur in both methods has not yet been made.

In this paper such a detailed correspondence between the tensor representations is given, and in a later paper it will be shown how the boson calculus can also be used for constructing spinor representations of all the orthogonal groups (not just the lower order groups which have a classical group as a covering group).

The boson calculus is usually concerned with the construction of unitary representations of unitary groups (including the unitary orthogonal and symplectic groups) whereas Zhelobenko considered the complex classical groups. However, many of the considerations involved in the construction of finite-dimensional representations are indifferent to whether a complex group or one of the real forms is being considered, and so long as questions of adjoint operators and scalar products do not arise, one may discuss representations without worrying about which particular field of numbers is chosen.

The starting point of both methods is to consider, say for the group $GL(n)$, polynomials in the matrix elements g_{ij}, and to define a representation by the right regular representation:

$$T_g f(g') = f(g'g). \tag{1}$$

Instead of the full group manifold of $GL(n)$ one may choose the homogeneous space $GL(n)/\mathrm{Z}D(n)$ where Z is a lower triangular group with unit diagonal elements and D is a diagonal matrix, and then the polynomial $f(g')$ is replaced by an inhomogeneous polynomial $f(z)$ where z_{ij} are the matrix elements of an upper triangular matrix, which serve to label the right cosets. Then multiplier representations of $GL(n)$ may be defined by right translations on this space according to

$$T_g f(z) = \alpha(zg) f(z \cdot g), \tag{2}$$

where $z \cdot g$ means the right component in the Gauss decomposition of the element $zg \in GL(n)$. This is the basis of Zhelobenko's method, and it is fully discussed in Ref. 2.

In its simplified form the boson calculus starts with an n-dimensional vector (a_i), $i = 1, \ldots, n$, which transforms according to the fundamental n-dimensional representation of $GL(n)$, homogeneous polynomials $f(a)$ in the (a_i) are constructed, and a representation is defined by

$$T_g f(a) = f(ag). \tag{3}$$

If we write $f(a) = a_1^m f(a_i/a_1)$, then $f(a_i/a_1)$ is a function over the homogeneous space $GL(n)/H$ where H is the subgroup for which $g_{1i} = 0, i > 1$, and in this simple case we see that the boson calculus is also concerned with a homogeneous space defined by a lower triangular subgroup. A more detailed examination given in the subsequent sections shows that the general boson calculus can be expressed in terms of polynomials over exactly the same homogeneous space as that employed by Zhelobenko.

In the next section the unitary and general linear groups will be discussed, and in the last section the orthogonal group.

2. THE UNITARY AND LINEAR GROUP

As already mentioned the boson calculus uses spaces of homogeneous polynomials to carry representations of the unitary group $U(n)$ and the general linear group. In order to relate the present discussion to the usual treatments of the boson calculus, $U(n)$ will be specifically referred to, with the proviso already made, that the results can be simply translated so as to apply to $GL(n)$.

In general we require n sets of bosons $a_i^\alpha (i, \alpha = 1, \ldots, n)$ with adjoints $\bar{a}_i^\alpha$ in order to obtain sufficient polynomials. It will be convenient to think of these operators as n^2 complex variables, i.e., $a_i^\alpha = z_i^\alpha \in C_{n^2}$, with adjoints $\bar{a}_i^\alpha = \partial/\partial z_i^\alpha$, and the vacuum state $|0\rangle$ becomes the constant 1. The representation space consists of polynomials homogeneous of degree m_k in the n variables a_i^k, for $k = 1, \ldots, n$. We can form an irreducible representation of $U(n)$ in the subspace of these polynomials in which the bosons a_i^α, for fixed α, appear only in antisymmetric combinations with $a_{i_1}^{\alpha-1}, a_{i_2}^{\alpha-2}, \ldots, a_i^1$. This subspace R_n appears through the application of the Young symmetrizer to an arbitrary polynomial, which may be regarded as a tensor under $U(n)$ transformations, to produce a polynomial (tensor) of a certain symmetry. The irreducible representation space now consists of polynomials homogeneous of degree $r_k = m_k - m_{k+1}$ in the variables $a_{i_1 i_2 \ldots i_k}$, for $k = 1, \ldots, n$ $(m_{n+1} = 0)$. The antisymmetric combination $a_{i_1 i_2 \ldots i_k}$ is defined by

$$a_{i_1 i_2 \ldots i_k} = \sum \epsilon(i_1 i_2 \ldots i_k) a_{i_1}^1 a_{i_2}^2 \ldots a_{i_k}^k, \tag{4}$$

which is the determinant of the $k \times k$ matrix $M_{ij} = a_j^i$. The numbers m_i, $i = 1, \ldots, n$ are nonnegative integers

satisfying $m_1 \geq m_2 \geq \cdots \geq m_n$ and label the representation of $U(n)$.

In this space R_n we define the irreducible representation T_g by (3) where "a" now stands collectively for the variables $a_{i_1 i_2 \ldots i_k}$, $k = 1, \ldots, n$ and ag stands for the same variables, in which each a_i^α has been transformed to $(a^\alpha g)_i = a_p^\alpha g_{pi}$ (summation over repeated $p, q, \ldots$). In this representation the generators of $U(n)$ have the form

$$E_{ij} = a_i^\dagger \bar{a}_j^p,$$

and satisfy

$$[E_{ij}, E_{kl}] = \delta_{jk} E_{il} - \delta_{il} E_{kj}.$$

The boson operators a_k^α behave as vectors under $U(n)$ transformations:

$$[E_{ij}, a_k^\alpha] = \delta_{jk} a_i^\alpha,$$

and the variables $a_{i_1 \ldots i_k}$ behave as tensors (multivectors) of rank k. The state of highest weight in the irreducible space R_n is (ignoring normalization)

$$|\max\rangle = a_1^{m_1 - m_2} a_{12}^{m_2 - m_3} \ldots a_{12\ldots n}^{m_n} |0\rangle. \tag{5}$$

By application of the diagonal generators $H_i = E_{ii}$ to this state we deduce that the numbers $m_1, m_2 \ldots m_n$ are the actual representation labels.

We wish to demonstrate the relation of these representations to those obtained by Zhelobenko[2] in a different formalism. In order to do this we will obtain another realization of the representation $T(g)$, in a projective space P_n which is set up in the following way (see, e.g., Hermann[3]). Two nonzero tensors a and a' are said to be equivalent if there is a nonzero scalar λ such that $a = \lambda a'$. P_n is then defined to be the set of all these equivalence classes, so that a point of P_n is an equivalence class of such tensors. A function f defined on P_n must then satisfy $f(\lambda a) = f(a)$, i.e., is homogeneous of zeroth degree. These functions can be constructed by taking functions in the inhomogeneous coordinates $a_{i_1 i_2 \ldots i_k} / a_{12\ldots k}$ for $k = 1, \ldots, n$. These coordinates are not defined everywhere, but on those points for which $a_{12\ldots k} \neq 0$. To each homogeneous polynomial defined in R_n there corresponds a single polynomial defined on P_n, since using the properties of $f \in R_n$ as a homogeneous polynomial we can write

$$f(a_i, a_{j_1 j_2}, \ldots, a_{i_1 \ldots i_n}) = a_1^{m_1 - m_2} a_{12}^{m_2 - m_3} \ldots a_{12\ldots n}^{m_n} \times f\left(\frac{a_i}{a_1}, \frac{a_{j_1 j_2}}{a_{12}}, \ldots, \frac{a_{i_1 \ldots i_n}}{a_{1\ldots n}}\right). \tag{6}$$

The functions on P_n are obtained by dividing the homogeneous polynomials by the state of highest weight (5). In this way the representation space can be characterized not as the space R_n of polynomials homogeneous in the variables a, but as the space of polynomials on the projective space P_n. This construction has been described before for $SU(2)$ (Vilenkin[4]). We have obtained the functions on P_n from the space R_n, in which the polynomial degrees m_k are also the representation labels. However, there are other ways to carry out the construction described, in which the degrees m_k are not the representation labels themselves.

The coordinate functions of P_n

$$\frac{a_{i_1 \ldots i_k}}{a_{1\ldots k}}, \quad k = 1, \ldots, n,$$

are not independent, but an independent set may be taken as

$$z_{ij} = \frac{a_{12\ldots i-1\, j}}{a_{12\ldots i-1\, i}}, \quad i, j = 1, \ldots, n.$$

This follows from the identities

$$a_{12\ldots m} a_{i_1 i_2 \ldots i_m i_{m+1}} = \sum_{p=1}^{m+1} (-)^{m^2+1+p} a_{i_1 \ldots i_{p-1} i_{p+1} \ldots i_{m+1}} \times a_{1\ldots m i_p}. \tag{7}$$

Using these identities for $m = 1, \ldots, n-1$ successively, we see using (6) that each

$$\frac{a_{i_1 \ldots i_k}}{a_{1\ldots k}}$$

can be expressed in terms of

$$\frac{a_{12\ldots m-1\, i}}{a_{12\ldots m-1\, m}} \quad \text{for } m = 1, \ldots, k.$$

Equation (7) is proved by considering the following $2m + 1 \times 2m + 1$ determinant. The upper left $m \times m$ block has elements $M_{kj} = a_j^k$; the upper right $m \times m + 1$ block is zero; the lower left $m + 1 \times m$ block has elements $M_{kj} = a_j^k$; and the lower right $m + 1 \times m + 1$ block has elements $L_{kj} = a_{i_j}^k$. The value of this determinant is $a_{12\ldots m} a_{i_1 \ldots i_{m+1}}$ as is shown by carrying out a Laplacian expansion according to the first m and the last $m + 1$ rows (see Aitken[5]). Now replace the ith row by the ith row minus the $(i + m)$th row, for $i = 1, \ldots, m$. The determinant is unchanged, but the upper left $m \times m$ block is now zero, and the upper right $m + 1 \times m$ block has elements $L_{kj} = -a_{i_j}^k$. Again carry out a Laplacian expansion, and we obtain the right-hand side of (7).

The function

$$f\left(\frac{a_i}{a_1}, \frac{a_{j_1 j_2}}{a_{12}}, \ldots, \frac{a_{i_1 \ldots i_n}}{a_{1\ldots n}}\right)$$

on P_n can now be written as a function ϕ of the variables z_{ij}. Hence the correspondence (6) may now be written

$$f(a) = a_1^{m_1 - m_2} a_{12}^{m_2 - m_3} \ldots a_{12\ldots n}^{m_n} \phi(z), \tag{8}$$

where $f \in R_n$, and ϕ is defined on $z = (z_{ij}) \in P_n$, where z is upper triangular. The function ϕ_g which corresponds to $f(ag)$ is then given by

$$\begin{aligned} f(ag) &= a_1^{m_1 - m_2} \ldots a_{12\ldots n}^{m_n} \phi_g \\ &= (ag)_1^{m_1 - m_2} (ag)_{12}^{m_2 - m_3} \ldots (ag)_{12\ldots n}^{m_n} \phi\left(\frac{(ag)_{1\ldots j}}{(ag)_{1\ldots i}}\right) \\ &= a_1^{m_1 - m_2} a_{12}^{m_2 - m_3} \ldots a_{12\ldots n}^{m_n} \\ &\quad \times \left(\frac{(ag)_1}{a_1}\right)^{m_1 - m_2} \left(\frac{(ag)_{12}}{a_{12}}\right)^{m_2 - m_3} \ldots \\ &\quad \times \left(\frac{(ag)_{12\ldots n}}{a_{12\ldots n}}\right)^{m_n} \phi\left(\frac{(ag)_{1\ldots j}}{(ag)_{1\ldots i}}\right). \end{aligned}$$

Therefore, the irreducible representation T_g defined on the space of functions ϕ on P_n is given by

$$T_g \phi(z) = \left(\frac{(ag)_1}{a_1}\right)^{m_1 - m_2} \ldots \left(\frac{(ag)_{12\ldots n}}{a_{12\ldots n}}\right)^{m_n} \phi(\tilde{z}), \tag{9}$$

where $\tilde{z}$ is the matrix with elements

$$\tilde{z}_{ij} = \frac{(ag)_{12\ldots i-1\,j}}{a_{12\ldots i-1\,i}} \Big/ \frac{(ag)_{12\ldots i-1\,i}}{a_{12\ldots i-1\,i}}.$$

The factors $\frac{(ag)_{12\ldots k}}{a_{12\ldots k}}$ for $k = 1, \ldots, n$ are functions of z_{ij}, the explicit form of which is given by

$$\frac{(ag)_{12\ldots k}}{a_{12\ldots k}} = \Delta_k(zg) \tag{10}$$

where $\Delta_k(zg)$ is the minor formed from the first k rows and columns of the matrix zg. To prove this, consider the $k \times k$ determinant D_k with elements

$$D_{ij} = a_{12\ldots i-1\,q} g_{qj} = a_{12\ldots i-1\,i} z_{iq} g_{qj} \quad (q \text{ summed})$$

which has the value

$$D_k = a_1 a_{12} \ldots a_{12\ldots k} \Delta_k(zg). \tag{11}$$

We will show that the (i,j) element of D_k may be written as $a^i_q g_{qj} a_{12\ldots i-1}$, without changing the value of D_k. This is clearly true for $i = 1$ (with the convention that $a_{12\ldots i-1} = 1$ for $i = 1$). Suppose it is true for $i = 1, \ldots, m-1$. We carry out the following row operations on the mth row leaving the determinant unchanged. Firstly, note that by expanding the determinant $a_{12\ldots m-1\,q}$ down the mth column, we can write

$$a_{12\ldots m-1\,q} = a^m_q a_{12\ldots m-1} + \sum_{r=1}^{m-1} a^r_q C_r$$

for some coefficient C_r (depending on a^α_l, $l \neq q$). Therefore,

$$a_{12\ldots m-1\,q} g_{qj} = a^m_q g_{qj} a_{12\ldots m-1} + \sum_{r=1}^{m-1} a^r_q g_{qj} C_r.$$

Now replace the mth row of D_k by

$$(\text{the } m\text{th row}) - \sum_{r=1}^{m-1} \frac{C_r}{a_{12\ldots r-1}} \quad (\text{the } r\text{th row}).$$

The element D_{rj} (for $r \leq m-1$) is $a^r_q g_{qj} a_{12\ldots r-1}$, so that now the element D_{mj} is equal to $a^m_q g_{qj} a_{12\ldots m-1}$. By induction, and by bringing out the factor $a_{12\ldots i-1}$, for $i = 2\ldots k$, we find that D_k is equal to $a_1 a_{12} \ldots a_{12\ldots k-1}$ multiplied by the determinant with elements $a^i_q g_{qj}$ which is $(ag)_{12\ldots k}$. This proves (10). If $\Delta_{ij}(zg)$ is the minor obtained from $\Delta_i(zg)$ by substituting the column with the number j in place of the column with the number i, then the same proof shows that

$$\Delta_{ij}(zg) = \frac{(ag)_{12\ldots i-1\,j}}{a_{12\ldots i-1\,i}}.$$

The irreducible representation T_g can now be written

$$T_g \phi(z) = \Delta_1(zg)^{m_1 - m_2} \Delta_2(zg)^{m_2 - m_3} \ldots \Delta_n(zg)^{m_n} \phi(\tilde{z})$$

where now we may write $\tilde{z}_{ij} = \Delta_{ij}(zg)/\Delta_i(zg)$. In this form we can see that the representation T_g is the same as that obtained by Zhelobenko by a different method. The results he has obtained can be immediately applied to our case where the functions $\phi_{\frac{12\ldots j}{12\ldots i}}$ are defined on the space P_n, with coordinates $z_{ij} = \frac{a_{12\ldots j}}{a_{12\ldots i}}$. On the other hand, in the formalism of Zhelobenko, the functions ϕ are defined on Z, the subgroup of $GL(n)$ consisting of upper triangular matrices with elements z_{ij}. However, we can exhibit P_n as a homogeneous space of $GL(n)$, and identify P_n with the coset space $GL(n)/H$, where $H = ZD(n)$ is the subgroup of lower triangular matrices referred to in Sec. 1. Let us determine the isotropy subgroup H of $GL(n)$ at the point in P_n determined by the tensors, denoted a, with the values $a_{i_1\ldots i_k} = 0$ except $a_{12\ldots k} = 1$, for $k = 1, \ldots, n$. The matrix g leaves the point in P_n fixed if there exist nonzero scalars $\lambda = \lambda(k)$ such that $ag = \lambda a$. Firstly, we show that $a^m_j = 0$, $j > m$, and $a^m_m = 1$. Clearly this is true for $m = 1$. If it is true for $m = 1, 2, \ldots, k-1$, then, writing $a_{12\ldots k-1\,j}$ as a $k \times k$ determinant, we find

$$a_{12\ldots k-1\,j} = a^1_1 a^2_2 \ldots a^{k-1}_{k-1} a^k_j = a^k_j \begin{cases} = 0 & \text{for } j > k \\ = 1 & \text{for } j = k \end{cases}$$

The result follows by induction. Now the condition $ag = \lambda a$ implies the result $g_{mj} = 0$, $j > m$, with $g_{mm} \neq 0$. This is true for $m = 1$, because $(ag)_i = \lambda(1)\, a_i = a_p g_{pi} = g_{1i} = \lambda(1)\delta_{1i}$. Hence $g_{1i} = 0$, $i > 1$, and $g_{11} = \lambda(1)$ is nonzero. Suppose the result is true for $m = 1, 2, \ldots, k-1$. Then $ag = \lambda a$ means

$$(ag)_{12\ldots k-1\,j} = \lambda(k)\, a_{12\ldots k-1\,j} = 0 \quad \text{for } j > k,$$
$$= \lambda(k) \quad \text{for } j = k.$$

Writing $(ag)_{12\ldots k-1\,j}$ as a determinant, we see that elements above the diagonal are zero. Therefore,

$$(ag)_{12\ldots k-1\,j} = g_{11} g_{22} \ldots g_{k-1\,k-1} g_{kj} = 0, \quad j > k$$
$$= \lambda(k) \neq 0, \quad j = k.$$

Hence $g_{kj} = 0$ for $j > k$, and $g_{kk} \neq 0$. By induction then we have shown that H is the subgroup of $GL(n)$ consisting of lower triangular matrices and we can put $P_n = GL(n)/H$. This demonstrates the asserted identity of the two methods for $GL(n)$.

3. ORTHOGONAL AND SYMPLECTIC GROUPS

In order to consider the groups with a metric some changes are necessary. The method for the symplectic group $Sp(n)$ is the same as that for the orthogonal group $O(n)$, except that the metric is antisymmetric, so we will consider only $O(n)$. In order that $g \in O(n)$, we require $g\sigma g^t = \sigma$, for some σ which is symmetric. We will choose for σ the matrix with elements $\sigma_{ij} = \delta_{i,n+1-j}$. The advantage of this metric is that the generators K_{ij} $(= -K_{ji})$, which satisfy

$$[K_{ij}, K_{kl}] = \sigma_{jk} K_{il} + \sigma_{il} K_{jk} - \sigma_{ik} K_{jl} - \sigma_{jl} K_{ik}, \tag{12}$$

are in Cartan standard form, i.e., are either weight generators, or raising or lowering generators. The classification of these generators, in terms of the generators E_α corresponding to the root α is as follows:

$$H_i = K_{n+1-i,i}, \quad i = 1, \ldots, \nu,$$
$$E_{e_p + e_q} = K_{n+1-q,n+1-p}, \quad E_{-e_p - e_q} = K_{pq}, \tag{13}$$
$$E_{e_p - e_q} = K_{n+1-p,q},$$

where $p, q = 1, \ldots, \nu$ and $\nu = [n/2]$. In addition, for $O(2\nu+1)$ we have $E_{e_p} = K_{2\nu+2-p,\nu+1}$ and $E_{-e_p} = K_{\nu+1,p}$, for $p = 1, \ldots, \nu$.

As before, the representations of $O(n)$ are realized in a space H_n of homogeneous polynomials. In order to

ensure that these representations are irreducible, the variables $a_i^\alpha (i, \alpha = 1, \ldots n)$ on which the polynomials are defined must be constrained with the condition $a_p^\alpha a_{n+1-p}^\beta = 0$. This is in order that the tensors of the representation space are traceless, or from the same viewpoint, the polynomials which appear are harmonic between all variables. We will think of the a_i^α as modified bosons,[1] i.e.,

$$a_i^\alpha = z_i^\alpha - 2\sigma(z^p, z^q)\Delta^{-1}_{(pq)(\alpha r)}\,\sigma_{is}\frac{\partial}{\partial z_s^r}$$

with adjoints $\bar{a}_i^\alpha = \partial/\partial z_i^\alpha$, but there are other possibilities. The traceless condition imposes the following conditions on the antisymmetric variables $a_{i_1\ldots i_k}, k = 1, \ldots, n$ [in the restriction to $SO(n)$ we have $k = 1, \ldots, \nu$]

$$a_{i_1\ldots i_k q}a_{j_1\ldots j_l\, n+1-q} = 0, \tag{14}$$

for if we expand according to Eq. (4) we see that each term has the factor $a_q^\alpha a_{n+1-q}^\beta$ (for some α, β) which is zero. The variables of our space will be $a_{i_1\ldots i_k}$ for $k = 1, \ldots, \nu$ and the polynomials will be homogeneous of degree $m_k - m_{k+1}$ in $a_{i_1\ldots i_k}(m_{\nu+1} = 0)$.

In this space H_n we define the irreducible representation T_g by $T_g f(a) = f(ag)$, $g \in O(n)$, $f \in H_n$. The generators then have the form

$$K_{ij} = \sigma_{iq}a_q^p\bar{a}_j^p - \sigma_{jq}a_q^p\bar{a}_i^p,$$

and the a_k^α behave as vectors under these transformations:

$$[K_{ij}, a_k^\alpha] = \delta_{jk}\sigma_{iq}a_q^\alpha - \delta_{ik}\sigma_{jq}a_q^\alpha.$$

For $O(2\nu + 1)$ the state of highest weight is

$$|\max\rangle = a_1^{m_1-m_2}a_{12}^{m_2-m_3}\ldots a_{12\ldots\nu}^{m_\nu}|0\rangle \tag{15}$$

and for $O(2\nu)$ there are two possibilities

$$|\max\rangle = a_1^{m_1-m_2}a_{12}^{m_2-m_3}\ldots a_{12\ldots\nu}^{m_\nu}|0\rangle \quad \text{for } m_\nu \geq 0 \tag{15'}$$

or

$$a_1^{m_1-m_2}\ldots a_{12\ldots\nu-1}^{m_{\nu-1}+m_\nu}a_{12\ldots\nu-1\,\nu+1}^{-m_\nu}|0\rangle \quad \text{for } m_\nu \leq 0.$$

These two cases arise because $a_{12\ldots\nu}a_{12\ldots\nu-1\,\nu+1} = 0$, a relation which follows easily from (14). When $m_\nu \leq 0$ the polynomials in H_n are homogeneous of degree $m_{\nu-1} + m_\nu$ in the $a_{i_1\ldots i_{\nu-1}}$ and of degree $-m_\nu$ in the $a_{i_1\ldots i_\nu}$. However, we will rarely consider the case $m_\nu \leq 0$ since the situation is entirely analogous to that for $m_\nu \geq 0$. The state of highest weight (15) has the same appearance as that for the unitary group (5). In fact, (15) is also the state of highest weight of the group $U(\nu)$ generated by $E^{\alpha\beta} = a_q^\alpha\bar{a}_q^\beta$, which commutes with $O(n)$. In the state ate (15) only the boson parts of the modified bosons give any contribution.

Now we define a projective space P_n in the same way as before, i.e., a point of P_n is an equivalence class of tensors, where two tensors a, a' are defined to be equivalent if $a = \lambda a'$ for a nonzero scalar λ. Functions on P_n are constructed from the coordinates $z_{ij} = a_{12\ldots i-1\,j}/a_{12\ldots i-1\,i}$. The a's here are constructed from modified bosons and will be manipulated formally, but the polynomials which actually appear are well defined as before because the variables $a_{12\ldots k}, k = 1, \ldots, \nu$ appearing in (15) may be regarded as ordinary bosons. The polynomials ϕ on P_n are obtained from those in H_n according to

$$f(a) = a_1^{m_1-m_2}\ldots a_{12\ldots\nu}^{m_\nu}\phi(z),$$

for the case $m_\nu \geq 0$. The z_{ij} are not all independent, and the relations between these variables are expressed in the restriction that z is orthogonal, i.e. $z\sigma z^t = \sigma$. Now $z_{i\,n+1-q}z_{jq} = \delta_{i\,n+1-j}$ holds identically for $i \geq n + 1 - j$ because z is upper triangular, with diagonal elements equal to one. For $i < n + 1 - j$ we need to show

$$a_{1\ldots i-1\;n+1-q}a_{1\ldots j-1\,q} = 0,$$

and this immediately follows from (14). These relations between the z_{ij} are the same as those between the z_{ij} which appear in the formalism described by Zhelobenko. The representations obtained in both approaches are therefore the same, with the same representation spaces, i.e., we have

$$T_g\,\phi(z) = \Delta_1(zg)^{m_1-m_2}\ldots\,\Delta_\nu(zg)^{m_\nu}\phi(\tilde{z}) \tag{16}$$

where in our formalism

$$\Delta_k(zg) = \frac{(ag)_{12\ldots k}}{a_{12\ldots k}}, \qquad k = 1, \ldots, \nu.$$

We can now use the important results obtained by Zhelobenko and apply them to our case. Of prime interest is the fact that the spinor representations fall naturally into this construction of the representations of $O(n)$. In the space of homogeneous polynomials H_n the numbers $m_1, m_2 \ldots m_\nu$ which label the representation can be integers only, and we do not obtain the two-valued (spinor) representations of $O(n)$, in which the m_i are all half odd-integers. Zhelobenko has explained in detail how these representations appear naturally in the space constructed according to (16). We will see in a later paper how to transfer back from the space of polynomials defined on P_n, to the space H_n of harmonic homogeneous polynomials in such a way as to include both spinor and tensor representations together in a natural way.

[1] M. A. Lohe and C. A. Hurst, J. Math. Phys. 12, 1882 (1971).
[2] D. P. Zhelobenko, Russ. Math. Surv. **XVII**, 1 (1962).
[3] R. Hermann, *Lie Groups for Physicists* (Benjamin, New York, 1966), Chap. 2.
[4] N. Y. Vilenkin, *Special Functions and Theory of Group Representations* (A.M.S. Transl., Providence, R.I. 1968), Chap. III.
[5] A. C. Aitken, *Determinants and Matrices* (Oliver and Boyd, Edinburgh, 1958), Chap. IV.

Spinor representations of the orthogonal groups

M. A. Lohe

Department of Mathematical Physics, University of Adelaide, Adelaide, S. Australia 5001
(Received 30 May 1973)

It is shown how to construct all spinor representations of $O(n)$. For the low order groups, this is done in such a way as to obtain the representation space of the covering group. More generally all representations of $O(n)$ are constructed in spaces of harmonic homogeneous polynomials. This is achieved for the spinor representations by finding realizations of the Lie algebra of $O(n)$ which are new. Results are written out explicitly for $O(3)$.

1. INTRODUCTION

The representations of $O(n)$ may be labelled by numbers $m_1, m_2, \ldots m_\nu$ where $\nu = [n/2]$ is the rank of $O(n)$. The m_i can be integers in which case the representations are single-valued and are referred to as the tensor representations. It has been shown in a previous paper[1] how to construct the irreducible tensor representations in spaces of harmonic homogeneous polynomials defined on modified boson operations a_i^α which satisfy $g_{pq}a_p^\alpha a_q^\beta = 0$, for some metric g. Since each polynomial behaves as a tensor under $O(n)$ transformations we could regard the representation space as consisting of tensors which are traceless, due to the condition $g_{pq}a_p^\alpha a_q^\beta = 0$. In this construction the labels m_i are also the polynomial degrees in a_j^i, so that the m_i are necessarily integers.

However, there exist also the spinor (two-valued) representations in which all the m_i are semi-integers (i.e., half odd integers), and which appear because $O(n)$ is not simply connected. It is important to be able to include the spinor representations in any construction of the $O(n)$ representations because of their physical importance, particularly for quantum mechanics. Such a construction is usually carried out by considering not actually the orthogonal group but its covering group, as is the case for $SO(3)$ covered by $SU(2)$ (see for example Bargmann[2]). This approach can be used only for the low order groups, namely $SO(3), SO(4)$ covered by $SU(2) \times SU(2), SO(5)$ covered by $Sp(4)$ and $SO(6)$ covered by $SU(4)$. We show in Sec. 2 how by consideration of these covering groups we obtain certain operators satisfying the traceless conditions, which are different from modified bosons. By substituting these operators into the expressions for the $SO(n)$ basis states, we obtain states in the representation space of the covering group. In this way we include the spinor representations by obtaining the representations of actually the covering group, even though these representations are constructed by a global analysis. Some of the operators used in this construction satisfy simple triple commutation relations, and these are also discussed briefly.

In general $SO(n)$ is not covered by another of the classical groups and we must find other methods to include the spinor representations. It is known how to construct the fundamental spinor representation using Clifford algebras (see for example Boerner[3]), and so we could produce arbitrary spinor representations by taking the direct product of a tensor representation and the fundamental spinor representation and reducing the result. However, this method is cumbersome, and even for $SO(3)$ the operators which take us directly into the irreducible subspace are without simple properties.

In a recent paper[4] (referred to as Paper I) we have shown how to obtain a construction of the representations of $SO(n)$ which includes both spinor and tensor representations together in a natural way. This construction was described by Zhelobenko[5] using the method of inducing representations, but we have shown how to repeat the construction using as a starting point the formalism of the boson calculus. By transferring back to this formalism we show that the space of harmonic homogeneous polynomials is in fact a suitable space with which to carry the spinor representations. This is achieved by finding realizations of the Lie algebra of $SO(n)$ which are new; here modified bosons play an important part.

For example, in the case $n = 3$ we find that contrary to general belief the spherical harmonics $Y_{lm}(\theta, \phi)$ are suitable basis states with which to carry the semi-integer representations, and these are distinguished from the integer representations by the different form of the generators. Attempts to construct the spinor representations in spaces of harmonic functions have centered on the interpretation of the $Y_{lm}(\theta, \phi)$ for semi-integer values of the label l. As previously indicated this is not possible, the problem being that even if the state of highest weight is given a meaning for semi-integer l, the lowering generator J_- does not annihilate $Y_{l,-l}(\theta, \phi)$ which must necessarily be the case. Pandres[6] has attempted to meet this difficulty by noting that $Q = J_- Y_{l,-l}$ is orthogonal to all Y_{lm}, but this property is not sufficient to show that $Q = 0$. In a representation space where no scalar product has been defined it is still necessary that J_- annihilates the minimum state (see for example Miller[7]).

The method described in Sec. 3 for $SO(n)$ leads also to new realizations of the Lie algebra of $U(n)$. It is likely that these realizations will be important in obtaining infinite-dimensional representations of the non-compact groups, especially considering that our approach is, in the words of Zhelobenko,[5] "the theory of finite-dimensional representations from the infinite-dimensional point of view."

2. COVERING GROUPS

In this section we demonstrate the relation between the orthogonal groups (of low order) and their covering groups, on a global scale. We give explicitly the substitutions necessary to transfer from $SO(n)$ basis states to those of the covering group, and so we obtain the spinor representations in the form of the representations of the covering group. Let us consider first $SO(3)$. Suppose we have three commuting operators a_1, a_2, a_3 (with adjoints $\bar{a}_1, \bar{a}_2, \bar{a}_3$) which form a vector with respect to the generators K_{ij} of $SO(3)$, i.e.,

$$[K_{ij}, a_k] = \delta_{jk}\sigma_{ip}a_p - \delta_{ik}\sigma_{jp}a_p \quad \text{(for notation see I).} \tag{2.1}$$

We also require that these vector components satisfy the traceless condition $2a_1a_3 + a_2^2 = 0$. We can then form an irreducible representation space with the functions obtained by allowing the a_i to act on the vacuum

$|0\rangle$ (for which $K_{ij}|0\rangle = 0$). The basis functions then have the form

$$|{}^{l}_{m}\rangle = a_1^m a_2^{l-m}|0\rangle, \tag{2.2}$$

where l labels the representation, and $-l \leq m \leq l$.

If we put the a_i equal to modified bosons with $K_{ij} = \sigma_{ip}a_p\bar{a}_j - \sigma_{jp}a_p\bar{a}_i$ we recover the usual $SO(3)$ basis states, the solid spherical harmonics, and l must be an integer. However, another possibility is to put

$$a_1 = \alpha_1^2, \quad a_2 = \sqrt{2}\alpha_1\alpha_2, \quad a_3 = -\alpha_2^2, \tag{2.3}$$

where α_1, α_2 are ordinary bosons. These operators satisfy $2a_1a_3 + a_2^2 = 0$ and form an $SO(3)$ vector, with the generators K_{ij} being given by

$$K_{31} = \tfrac{1}{2}(\alpha_1\bar{\alpha}_1 - \alpha_2\bar{\alpha}_2), \quad K_{21} = (1/\sqrt{2})\alpha_2\bar{\alpha}_1, \quad K_{32} = (1/\sqrt{2})\alpha_1\bar{\alpha}_2. \tag{2.4}$$

We have in fact obtained the $SU(2)$ representation space. The basis states (2.2) become

$$|{}^{l}_{m}\rangle = \alpha_1^{l+m}\alpha_2^{l-m}|0\rangle$$

and l can be a semi-integer. The operators (2.3) can be obtained by comparing states of highest weight for $SO(3)$ and $SU(2)$, after equating the generators for each as in (2.4). We obtain $a_1 = \alpha_1^2$, and then by requiring that (2.1) should hold we deduce the expressions for a_2 and a_3. The operators α_1, α_2 are components of a vector under $SU(2)$, but are regarded as spinor components under $SO(3)$.

We can use the same technique for $SO(4)$. The generators of $SU(2) \times SU(2)$ may be expressed with $E_{ij} = \alpha_i\bar{\alpha}_j$, $i,j = 1,\ldots,4$, where the α's are ordinary bosons, and then the correspondence between the generators of $SO(4)$ and $SU(2) \times SU(2)$ is

$\frac{1}{2}(K_{41} + K_{32})$	$\frac{1}{2}(E_{11} - E_{22})$
$\frac{1}{2}(K_{41} - K_{32})$	$\frac{1}{2}(E_{33} - E_{44})$
K_{43}	E_{12}
K_{21}	E_{21}
K_{42}	E_{34}
K_{31}	E_{43}

(2.5)

The state of highest weight is then

$$|\max\rangle = a_1^{m_1-m_2}a_{12}^{m_2}|0\rangle, \quad m_2 \geq 0 = \alpha_1^{m_1+m_2}\alpha_3^{m_1-m_2}|0\rangle,$$

so that $a_1 = \alpha_1\alpha_3$, $a_{12} = \alpha_1^2$. Now from the requirement that $a_1 = \alpha_1\alpha_3$ behaves like the first component of an $SO(4)$ vector under the E_{ij}, we find

$$a_1 = \alpha_1\alpha_3, \quad a_2 = \alpha_1\alpha_4, \quad a_3 = \alpha_2\alpha_3, \quad a_4 = -\alpha_2\alpha_4. \tag{2.6}$$

Similarly, we obtain all $a_{ij} = a_ib_j - a_jb_i$, so that we can solve for all b_i. It is possible only to express the b's in terms of any one of them, b_1 say. We find

$$b_2 = \frac{\alpha_1}{\alpha_3} + \frac{\alpha_4}{\alpha_3}b_1, \quad b_3 = \frac{\alpha_2}{\alpha_1}b_1, \quad b_4 = -\frac{\alpha_2\alpha_4}{\alpha_1\alpha_3}b_1 - \frac{\alpha_2}{\alpha_3}. \tag{2.7}$$

The variables a_{ij} which actually appear are nevertheless independent of b_1. The a's and b's found in this way satisfy the traceless conditions, i.e., $a_1a_4 + a_2a_3 = 0$, $b_1b_4 + b_2b_3 = 0$, $a_1b_4 + a_2b_3 + a_3b_2 + a_4b_1 = 0$. In fact they satisfy more than this, because we have $a_1b_4 + a_3b_2 = 0 = a_4b_1 + a_2b_3$. This is because a_{13} and a_{24} are zero in the space for which $m_2 > 0$. When $m_2 < 0$ the expressions for the b's are

$$b_2 = \frac{\alpha_4}{\alpha_3}b_1, \quad b_3 = \frac{\alpha_2}{\alpha_1}b_1 + \frac{\alpha_3}{\alpha_1}, \quad b_4 = -\frac{\alpha_2\alpha_4}{\alpha_1\alpha_3}b_1 - \frac{\alpha_4}{\alpha_1}. \tag{2.8}$$

and then $a_{12} = 0 = a_{34}$.

The expressions for the a_i as double bosons satisfy the following simple triple commutation relations:

$$[a_i, a_j] = 0, \quad [a_i, [\bar{a}_j, a_k]] = \sigma_{ik}\sigma_{jp}a_p - \delta_{jk}a_i - \delta_{ij}a_k \quad i,j,k = 1,\ldots,n. \tag{2.9}$$

The operators (2.3) for $SO(3)$ also satisfy these relations, but the b_i (2.7) and (2.8) do not have this property.

It is of interest to find other operators satisfying (2.9) because of the possibility that there could be other solutions of the traceless conditions. However, it is possible to classify all operators expressed as double bosons which satisfy (2.9), and it can then be shown that (2.3) and (2.6) are essentially the only solutions with $\sigma_{pq}a_pa_q = 0$. However, for arbitrary n the operators

$$a_i = \alpha_i\sqrt{n/2 + N - 1}$$

satisfy (2.9), where α_i are modified bosons defined by

$$[\alpha_i, \alpha_j] = 0, \quad [\bar{\alpha}_i, \alpha_j] = \delta_{ij} - \sigma_{ip}\alpha_p(n/2 + N)^{-1}\sigma_{jq}\bar{\alpha}_q,$$

where $N = \alpha_p\bar{\alpha}_p$ is the number operator. If we put $a_i = \alpha_i\sqrt{p + N}$ where the α_i are ordinary bosons we obtain commutation relations similar to (2.9), namely

$$[a_i, a_j] = 0, \quad [a_i, [\bar{a}_j, a_k]] = -\delta_{jk}a_i - \delta_{ij}a_k. \tag{2.10}$$

For the case $n = 1$ these relations are the same as (2.9) and this case has been studied by Kademova and Kraev.[8] The algebra of the a_i is isomorphic to the Lie algebra of $SU(n,1)$[9] for (2.10) and isomorphic to the Lie algebra of $O(n,2)$ for (2.9).

The techniques we have used for $SO(3)$ and $SO(4)$ to find operators which permit the construction of all the representations of the covering group apply also to $SO(5)$ and $SO(6)$. For $SO(5)$ a_i, b_i, $i = 1,\ldots,5$, are expressed in terms of modified bosons α_i, β_i for $Sp(4)$ which satisfy $\epsilon_{pq}\alpha_p\beta_q = 0$. Putting

$$\epsilon = \begin{pmatrix} 0 & -1 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & -1 \\ 0 & 0 & 1 & 0 \end{pmatrix},$$

we have $\alpha_{12} + \alpha_{34} = 0$. The correspondence between the generators of $Sp(4)$ and $SO(5)$ is as follows, where

$$S_{ij} = \epsilon_{ip}\alpha_p\bar{\alpha}_j + \epsilon_{jp}\alpha_p\bar{\alpha}_i + \epsilon_{ip}\beta_p\bar{\beta}_j + \epsilon_{jp}\beta_p\bar{\beta}_i, \qquad i,j = 1,\ldots,4,$$

are the generators of $Sp(4)$:

S_{21}	$K_{51} + K_{42}$
S_{43}	$K_{51} - K_{42}$
$1/\sqrt{2}S_{42}$	K_{53}
$1/\sqrt{2}S_{32}$	K_{43}
$-1/\sqrt{2}S_{31}$	K_{31}
$1/\sqrt{2}S_{41}$	K_{32}
$\frac{1}{2}S_{22}$	K_{54}
$-\frac{1}{2}S_{44}$	K_{52}
$-\frac{1}{2}S_{11}$	K_{21}
$\frac{1}{2}S_{33}$	K_{41}

The operators a_i, b_i which have all the required properties are $a_1 = \alpha_{13}$, $a_2 = -\alpha_{14}$, $a_3 = \sqrt{2}\alpha_{12}$, $a_4 = \alpha_{23}$, $a_5 = \alpha_{24}$,

$$b_2 = \frac{\alpha_1^2}{\alpha_{13}} - \frac{\alpha_{14}}{\alpha_{13}}b_1,$$

$$b_3 = \sqrt{2}\,\frac{\alpha_1\alpha_3}{\alpha_{13}} + \sqrt{2}\,\frac{\alpha_{12}}{\alpha_{13}}b_1,$$

$$b_4 = -\frac{\alpha_3^2}{\alpha_{13}} + \frac{\alpha_{23}}{\alpha_{13}}b_1,$$

$$b_5 = -\frac{\alpha_1\alpha_2}{\alpha_{13}} - \frac{\alpha_3\alpha_4}{\alpha_{13}} + \frac{\alpha_{24}}{\alpha_{13}}b_1. \qquad (2.12)$$

For $SO(6)$ the operators a_i, b_i, c_i, $i = 1,\ldots,6$, are expressed in terms of ordinary bosons. The results for a_i are $a_1 = \alpha_{12}$, $a_2 = \alpha_{13}$, $a_3 = \alpha_{23}$, $a_4 = \alpha_{14}$, $a_5 = -\alpha_{24}$, $a_6 = \alpha_{34}$.

3. REPRESENTATIONS IN SPACES OF HARMONIC HOMOGENEOUS POLYNOMIALS

It is necessary to use more general methods in order to construct both spinor and tensor representations of $SO(n)$, for arbitrary dimension n. We wish to develop the boson calculus further to allow this, i.e., to construct these representations in spaces H_n of harmonic homogeneous polynomials. A unified construction has been carried out by Zhelobenko[5] by defining multiplier representations in the space of functions on a subgroup Z consisting of upper triangular matrices. In I (to which we refer for notation and further details) it has been shown how to construct these same mulitplier representations in the space of functions ϕ on a set of $n \times n$ upper triangular matrices z with elements z_{ij}. The functions ϕ, homogeneous of zeroth degree, are constructed by taking rational functions in modified boson operators $a_i^\alpha(i,\alpha = 1,\ldots,n)$, and the coordinate functions z_{ij} can be put equal to $a_{1\ldots i-1\,j}/a_{1\ldots i-1\,i}$. If $f(a)$ is a harmonic polynomial, homogeneous of degree $r_k = m_k - m_{k+1}$ in $a_{i_1\ldots i_k}$ (for $k = 1,\ldots,\nu$, $m_{\nu+1} = 0$), then the 1-1 correspondence between $f(a)$ and $\phi(z)$ is given by

$$f(a) = a_1^{m_1-m_2}\ldots a_{12\ldots\nu-1}^{m_{\nu-1}-m_\nu}a_{12\ldots\nu}^{m_\nu}\phi(z). \qquad (3.1)$$

The representation T_g in the space of functions ϕ is given by

$$T_g\phi(z) = \Delta_1^{m_1-m_2}\ldots\Delta_\nu^{m_\nu}\phi(\tilde{z}), \qquad (3.2)$$

where $\Delta_k(z,g)$ is the minor formed from the first k rows and columns of the matrix zg and is equal to $(ag)_{12\ldots k}/a_{12\ldots k}$, and $\tilde{z}$ is a known function of z. The important results which Zhelobenko has obtained and which apply to T_g defined by (3.2) are as follows:

For $n = 2\nu + 1$ there exists a polynomial $\mathfrak{S}_0(z,g)$ on P_n such that $\Delta_\nu(zg) = \mathfrak{S}_0^2(z,g)$. The fundamental spinor representation labelled by $(\frac{1}{2},\frac{1}{2}\ldots\frac{1}{2})$ can now be constructed according to $S_g\phi(z) = \mathfrak{S}_0(z,g)\phi(\tilde{z})$, where S_g is a representation in the sense that $S_{g_1g_2} = \pm S_{g_1}S_{g_2}$. The multiplier for an arbitrary representation T_g is now written

$$\Delta_1^{m_1-m_2}\ldots\Delta_{\nu-1}^{m_{\nu-1}-m_\nu}\mathfrak{S}_0^{2m_\nu}, \qquad (3.3)$$

so that m_ν may be a semi-integer, in which case $m_1,\ldots,m_{\nu-1}$ are also semi-integers.

For $n = 2\nu$ there exist two polynomials on P_n, $\mathfrak{S}_-$ and $\mathfrak{S}_+$ such that

$$\Delta_{\nu-1}(zg) = \mathfrak{S}_-(z,g)\,\mathfrak{S}_+(z,g),$$
$$\Delta_\nu(zg) = \mathfrak{S}_+^2(z,g).$$

The two fundamental spinor representations, labelled by $(\frac{1}{2},\frac{1}{2},\ldots,\pm\frac{1}{2})$ are constructed according to $S_g\phi(z) = \mathfrak{S}_\pm\phi(\tilde{z})$ and in general the multiplier has the form

$$\Delta_1^{m_1-m_2}\ldots\Delta_{\nu-2}^{m_{\nu-2}-m_{\nu-1}}\mathfrak{S}_-^{m_{\nu-1}-m_\nu}\mathfrak{S}_+^{m_{\nu-1}+m_\nu}. \qquad (3.4)$$

Again $m_\nu, m_{\nu-1}$ can be semi-integers simultaneously, in which case $m_1,\ldots,m_{\nu-2}$ are all semi-integers. In the form (3.4) the representation includes naturally the case for which $m_\nu < 0$, even though the transfer (3.1) to the functions ϕ has been carried out from polynomials f for which $m_\nu \geq 0$ only. If we had begun with functions f for which $m_\nu < 0$, then we would reach the same space of functions $\phi(z)$ by putting

$$f(a) = a_1^{m_1-m_2}\ldots a_{12\ldots\nu-1}^{m_{\nu-1}+m_\nu}a_{1\ldots\nu-1\,\nu+1}^{-m_\nu}\phi(z)$$

with

$$z_{\nu j} = \frac{a_{12\ldots\nu-1\,j}}{a_{12\ldots\nu-1\,\nu+1}}, \qquad j \geq \nu + 2$$

($z_{\nu\,\nu+1} = a_{1\ldots\nu-1\,\nu}/a_{1\ldots\nu-1\,\nu+1} = 0$ as before).

We wish to transfer back to the space H_n in such a way as to retain this construction of the spinor representations. We obtain a polynomial $f(a)$ in H_n from $\phi(z)$ by multiplying ϕ with a certain polynomial which becomes the state of highest weight in H_n. This is expressed in the formula (3.1) which holds in the case when the degrees r_i of $f(a) \in H_n$ are connected with the representation labels m_i by $r_i = m_i - m_{i+1}$. As previously noted, the formula (3.1) restricts each m_i to nonnegative integral values. More generally, however, we can also transfer back to H_n by multiplying each $\phi(z)$ with a polynomial of degrees r_i such that $r_i \geq m_i - m_{i+1}$. We do this in the following way. In the representation space of polynomials $\phi(z)$ on P_n we replace each m_i by $m_i - p_i = l_i$, for $i = 1,\ldots,\nu$, so that the representation labels are now $l_i = m_i - p_i$. Now transfer back to H_n, the space of harmonic polynomials in the a's with degrees r_i where r_i is not equal to $l_i - l_{i+1} = (m_i - m_{i+1}) - (p_i - p_{i+1})$ as before, but $r_i = m_i - m_{i+1}$. This transfer is carried out by multi-

plying each $\phi(z)$ with the polynomial $a_1^{m_1-m_2}\dots a_{12\dots\nu}^{m_\nu}$, i.e., (3.1) still holds, but now the m_i are no longer the representation labels. We can be sure that $f(a)$ obtained according to (3.1) is actually a polynomial in the a's if $m_i - m_{i+1} \geq (m_i - p_i) - (m_{i+1} - p_{i+1})$ for $i = 1, \dots, \nu (p_{\nu+1} = 0)$, i.e., if $p_1 \geq p_2 \geq \dots \geq p_\nu \geq 0$. Since the $m_i - p_i$ are representation labels they satisfy

$$m_1 - p_1 \geq m_2 - p_2 \geq \dots \geq m_\nu - p_\nu \geq 0, \quad n = 2\nu + 1,$$

and

$$m_1 - p_1 \geq m_2 - p_2 \geq \dots \geq |m_\nu - p_\nu|, \quad n = 2\nu. \tag{3.5}$$

The m_i are integers, hence the p_i are either all integers or all semi-integers.

In this way we obtain the representations $(l_1, \dots, l_\nu)$ in the space of harmonic polynomials homogeneous of degrees $r_i = m_i - m_{i+1}$. By choosing p_i suitably we can obtain any of the permissible values of $(l_1, \dots, l_\nu)$. All the tensor representations for $l_\nu \geq 0$ are obtained by putting $p_i = 0$ for all i, and all the spinor representations for $l_\nu \geq 0$ by putting $p_i = \frac{1}{2}$ for all i. We could also obtain, for $n = 2\nu$, all representations for which $l_\nu < 0$ but these are constructed more conveniently in the space H_n for which $m_\nu < 0$.

The representation T_g in the space of functions $\phi(z)$ has the form (putting $m_i \to m_i - p_i$)

$$T_g\phi(z) = \{\Delta_1^{m_1-m_2}\dots\Delta_\nu^{m_\nu}\phi(\tilde{z})\}\left(\frac{1}{\Delta_1}\right)^{p_1-p_2}\dots\left(\frac{1}{\Delta_\nu}\right)^{p_\nu} = \phi_g\left[\frac{1}{\Delta_1}\right]^{p_1-p_2}\dots\left[\frac{1}{\Delta_\nu}\right]^{p_\nu}.$$

We transfer back to H_n and using the fact that ϕ_g corresponds to $f(ag)$ and $\Delta_k = (ag)_{12\dots k}/a_{12\dots k}$ we find that T_g is defined in H_n by

$$T_g f(a) = \left(\frac{a_1}{(ag)_1}\right)^{p_1-p_2}\dots\left(\frac{a_{12\dots\nu}}{(ag)_{12\dots\nu}}\right)^{p_\nu} f(ag) \tag{3.6}$$

For $n = 2\nu + 1$ this may be written

$$T_g f(a) = \left(\frac{a_1}{(ag)_1}\right)^{p_1-p_2}\dots\left(\frac{a_{12\dots\nu-1}}{(ag)_{12\dots\nu-1}}\right)^{p_{\nu-1}-p_\nu}\mathfrak{S}_0^{-2p_\nu} f(ag), \tag{3.7}$$

while for $n = 2\nu$ we have

$$T_g f(a) = \left(\frac{a_1}{(ag)_1}\right)^{p_1-p_2}\dots\left(\frac{a_{12\dots\nu-2}}{(ag)_{12\dots\nu-2}}\right)^{p_{\nu-2}-p_{\nu-1}}\mathfrak{S}_-^{p_\nu-p_{\nu-1}}\dots\mathfrak{S}_+^{-p_{\nu-1}-p_\nu} f(ag), \tag{3.8}$$

where $\mathfrak{S}_0, \mathfrak{S}_\pm$ are polynomials in $a_{12\dots i-1\,j}/a_{12\dots i-1\,i}$.

The space H_n is invariant under T_g provided the parameters p_i are restricted to the values indicated above. From T_g we can calculate the form of the generators. However, we will find it easier to use the correspondence (3.1) to calculate the dependence of the generators on m_i in the space of functions ϕ, and then to put $m_i \to m_i - p_i$ and transfer back to H_n. If

$$K_{ij}(a) = \sigma_{iq}a_q^p\bar{a}_j^p - \sigma_{jp}a_q^p\bar{a}_i^p,$$

then the generators $K_{ij}(z)$ acting on $\phi(z)$ are determined by

$$K_{ij}(a)f(a) = K_{ij}(a)a_1^{m_1-m_2}\dots a_{12\dots\nu}^{m_\nu}\phi\left(\frac{a_{1\dots k-1\,l}}{a_{1\dots k-1\,k}}\right) = a_1^{m_1-m_2}\dots a_{12\dots\nu}^{m_\nu}K_{ij}(z)\phi(z). \tag{3.9}$$

The dependence of $K_{ij}(z)$ on the m_i, which is found from the action of $K_{ij}(a)$ on $a_1^{m_1-m_2}\dots a_{12\dots\nu}^{m_\nu}$, is established in this way for integral m_i but will also hold in the case when m_i takes semi-integer values. The classification of the K_{ij} as raising or lowering generators, or weight generators, has been given in I. The raising generators commute with $a_1^{m_1-m_2}\dots a_{12\dots\nu}^{m_\nu}$ and therefore when acting on $\phi(z)$ are independent of m_i. Hence they are unchanged in H_n. The weight generators are $H_i = K_{n+1-i,\,i}$ and using (3.9) we see that $K_{n+1-i,\,i}(z) = m_i + D_i(z)$ where D_i is a differential operator in z, independent of $m_1, \dots, m_\nu$. Putting $m_i \to m_i - p_i$ and transferring to H_n, we have that $K_{n+1-i,\,i}(a)$ is replaced by $K_{n+1-i,\,i}(a) - p_i$. In order to specify the changes necessary for the lowering generators, it is sufficient to consider only the generators corresponding to the simple roots, since all other lowering generators are obtained from these by commutation. For $n = 2\nu + 1$ the lowering generators corresponding to the simple roots are $K_{2\nu+1-j,\,j}$ for $j = 1, \dots, \nu$. We find $K_{2\nu+1-j,\,j}(z) = (m_j - m_{j+1})z_{j\,j+1} + D_j'(z)$ for some D'. Hence in H_n $K_{2\nu+1-j,\,j}(a)$ is replaced by

$$K_{2\nu+1-j,\,j}(a) - (p_j - p_{j+1})\frac{a_{1\dots j-1\,j+1}}{a_{1\dots j-1\,j}}, \quad j = 1, \dots \nu.$$

For $n = 2\nu$ all lowering generators can be obtained from $K_{2\nu,\,j}$ $(j = 1, \dots, \nu - 1)$ and $K_{\nu,\,\nu-1}$. We find $K_{2\nu-j,\,j}(z) = (m_j - m_{j+1})z_{j\,j+1} + D_j''(z)$ and $K_{\nu,\,\nu-1}(z) = (m_{\nu-1} + m_\nu)z_{\nu-1\,\nu+1} + D_j'''(z)$ for differential operators D'', D'''. Hence in H_n

$$K_{2\nu-j,\,j}(a) \to K_{2\nu-j,\,j}(a) - (p_j - p_{j+1})\frac{a_{1\dots j-1\,j+1}}{a_{1\dots j-1\,j}}$$

and

$$K_{\nu,\,\nu-1}(a) \to K_{\nu,\,\nu-1}(a) - (p_{\nu-1} + p_\nu)\frac{a_{1\dots\nu-2\,\nu+1}}{a_{1\dots\nu-2\,\nu-1}}.$$

These replacements are considerably simplified in the case $p_i = 0$ (tensor representations) and $p_i = \frac{1}{2}$ (spinor representations). Although the generators involve ratios of the variables $a_{i_1\dots i_k}$ their range is a subspace of H_n and no rational function of polynomials appears provided the p_i satisfy (3.5) and $p_1 \geq p_2 \geq \dots \geq p_\nu \geq 0$. The representations constructed are not unitary in general, although they are equivalent to unitary representations. They can be made unitary by redefining the scalar product in H_n which can always be done because $SO(n)$ is compact (see Vilenkin, Ref. 10, p. 44).

It is possible to find realizations of the type just described for the generators of $U(n)$ also. We can construct representations labelled by $l_i = m_i - p_i$ $(i = 1, \dots, n)$ in the space of homogeneous polynomials of degree $r_k = m_k - m_{k+1}$ in the variables $a_{i_1\dots i_k}$ for $k = 1, \dots, n$, where a_i^α are now ordinary bosons. We require $p_1 \geq p_2 \dots \geq p_n \geq 0$ where the p_i are all integers, and also $m_1 - p_1 \geq m_2 - p_2 \dots \geq m_n - p_n$. The representa-

tion T_g in this space is given by

$$T_g f(a) = \left(\frac{a_1}{(ag)_1}\right)^{p_1-p_2} \cdots \left(\frac{a_{12\ldots n}}{(ag)_{12\ldots n}}\right)^{p_n} f(ag) \qquad (3.10)$$

and the generators E_{ij} satisfying

$$[E_{ij}, E_{kl}] = \delta_{jk}E_{il} - \delta_{il}E_{kj}$$

are specified by

$$\begin{aligned} &E_{ij} = a_i^p \bar{a}_j^p, \quad j > i \text{ (raising generators)},\\ &H_i = E_{ii} = a_i^p \bar{a}_i^p - p_i,\\ &E_{i+1,i} = a_{i+1}^p \bar{a}_i^p - (p_i - p_{i+1}) \frac{a_{1\ldots i-1\, i+1}}{a_{1\ldots i-1\, i}}\\ &\qquad\qquad \text{(lowering generators)}. \end{aligned} \qquad (3.11)$$

This construction leads to no new representations except that now l_n, the last label, can be negative in addition to the usual nonnegative values.

4. RESULTS FOR $O(3)$

In order to illustrate the construction described above we will write down the results explicitly for $SO(3)$. We begin in the space H of harmonic homogeneous polynomials of degree m_1 in the a_i where $m_1 = l$ is also the representation label, and the a_i are modified bosons with $2a_1a_3 + a_2^2 = 0$. We have

$$\begin{aligned} J_+ &= K_{32} = a_1\bar{a}_2 - a_2\bar{a}_3,\\ J_- &= K_{21} = a_2\bar{a}_1 - a_3\bar{a}_2,\\ J_3 &= K_{31} = a_1\bar{a}_1 - a_3\bar{a}_3, \end{aligned}$$

and an arbitrary basis state is

$$|{}^l_m\rangle = a_1^m a_2^{l-m}|0\rangle.$$

We put $z_{12} = a_2/a_1 = z$, so that $a_3/a_1 = -\frac{1}{2}z^2$ and then

$$\begin{aligned} f(a_1, a_2, a_3) &= a_1^{m_1} f\left(1, \frac{a_2}{a_1}, \frac{a_3}{a_1}\right)\\ &= a_1^{m_1}\phi(z), \end{aligned}$$

where

$$\phi(z) = f(1, z, -\tfrac{1}{2}z^2).$$

The representation T_g in the space of functions $\phi(z)$ is given by

$$T_g\phi(z) = \left(\frac{(ag)_1}{a_1}\right)^{m_1} \phi\left(\frac{(ag)_2}{a_1} \Big/ \frac{(ag)_1}{a_1}\right).$$

Now $g = (g_{ij})$ satisfies $g\sigma g^t = \sigma$, so that if g_{11}, g_{12}, g_{21} are taken to be independent, we have

$$\begin{aligned} g_{31} &= -\frac{g_{21}^2}{2g_{11}}, \quad g_{13} = -\frac{g_{12}^2}{2g_{11}},\\ g_{32} &= -\frac{g_{12}g_{21}^2}{2g_{11}^2} - \frac{g_{21}}{g_{11}}, \quad g_{23} = -\frac{g_{21}g_{12}^2}{2g_{11}^2} - \frac{g_{12}}{g_{11}},\\ g_{22} &= \frac{g_{21}g_{12}}{g_{11}} + 1, \quad g_{33} = \frac{1}{g_{11}}\left(1 + \frac{g_{21}g_{12}}{2g_{11}}\right)^2. \end{aligned} \qquad (4.1)$$

Hence

$$\begin{aligned} (ag)_1/a_1 &= \frac{a_p}{a_1} g_{p1} = z_{1p}g_{p1}\\ &= g_{11} + zg_{21} + \tfrac{1}{2}z^2 \cdot (g_{21}^2/2g_{11})\\ &= (\sqrt{g_{11}} + zg_{21}/2\sqrt{g_{11}})^2, \end{aligned}$$

i.e.,

$$\mathfrak{S}_0(z, g) = \sqrt{g_{11}} + zg_{21}/2\sqrt{g_{11}}.$$

Also,

$$\frac{(ag)_2}{a_1} = \left(\sqrt{g_{11}} + \frac{zg_{21}}{2\sqrt{g_{11}}}\right)\left(\frac{g_{12}}{\sqrt{g_{11}}} + \frac{z}{\sqrt{g_{11}}} + z\,\frac{g_{12}g_{21}}{2g_{11}\sqrt{g_{11}}}\right);$$

therefore

$$T_g\phi(z) = \left(\sqrt{g_{11}} + \frac{zg_{21}}{2\sqrt{g_{11}}}\right)^{2m_1} \phi\left(\frac{g_{12} + z + zg_{12}g_{21}/2g_{11}}{g_{11} + \frac{1}{2}zg_{21}}\right). \qquad (4.2)$$

The basis functions are z^{l-m} for $-l \leq m \leq l$ and the generators are

$$J_+ = \frac{d}{dz}, \quad J_- = m_1 z - \tfrac{1}{2}z^2\,\frac{d}{dz}, \quad J_3 = m_1 - z\,\frac{d}{dz}$$

m_1 can now be a semi-integer. Putting $m_1 \to m_1 - p = l$, where p can have any of the values $0, \frac{1}{2}, \ldots, m_1$ and transferring back to H, we find the generators have the form

$$\begin{aligned} J_+ &= a_1\bar{a}_2 - a_2\bar{a}_3,\\ J_- &= a_2\bar{a}_1 - a_3\bar{a}_2 - p\,\frac{a_2}{a_1}\\ J_3 &= a_1\bar{a}_1 - a_3\bar{a}_3 - p. \end{aligned} \qquad (4.3)$$

T_g is defined in H by

$$T_g f(a) = \left(\frac{2a_1\sqrt{g_{11}}}{2a_1g_{11} + a_2g_{21}}\right)^{2p} f(ag). \qquad (4.4)$$

We obtain all tensor representations by putting $p = 0$, and all spinor representations by putting $p = \frac{1}{2}$.

The basis states are

$$|{}^l_m\rangle = a_1^{m+p}a_2^{l-m}|0\rangle,$$

which are the usual spherical harmonics as can be seen by expanding the modified bosons

$$a_i = x_i - r^2\,\frac{1}{3 + 2N}\,\sigma_{ip}\,\frac{d}{dx_p},$$

where $r^2 = 2x_1x_3 + x_2^2$. This expansion is carried out most easily by writing

$$a_1^{m+p}a_2^{l-m}|0\rangle = Hx_1^{m+p}x_2^{l-m},$$

where H is a projection operator calculated by Vilenkin (Ref. 10, p. 446). Another expression for $|{}^l_m\rangle$ can be written by noting that

$$\begin{aligned} a_2h^l &= \left(x_2 - \frac{r^2}{2l+1}\,\frac{d}{dx_2}\right)h^l\\ &= -\frac{1}{2l+1}\,r^{2l+3}\,\frac{d}{dx_2}\,\frac{1}{r^{2l+1}}\,h^l, \end{aligned}$$

where $h^l \in H$ is of degree l. In this form a_2 appears

as the well-known Maxwell multipole. The basis states are now written as

$$|{}^{l}_{m}\rangle \propto r^{2l+2p+1} x_1^{m+p}\left(\frac{d}{dx_2}\right)^{l-m} \frac{1}{r^{2m+2p+1}}.$$

We can transform to spherical polar coordinates by putting

$$x_1 = \frac{r}{\sqrt{2}}\, e^{i\phi} \sin\theta,$$
$$x_2 = r\cos\theta,$$
$$x_3 = \frac{r}{\sqrt{2}}\, e^{-i\phi} \sin\theta,$$

and then

$$|{}^{l}_{m}\rangle \propto r^{l+p} Y_{l+p,\,m+p}(\theta,\phi).$$

The minimum state is

$$|{}^{l}_{-l}\rangle = a_2^{2p} a_3^{l-p}|0\rangle \text{ if } p \leq l,$$

and

$$|{}^{l}_{-l}\rangle = a_1^{p-1} a_2^{2l}|0\rangle \text{ if } p \geq l.$$

These states are an orthogonal basis, for although J_+ is not the Hermitian adjoint of J_- the labelling operators J_3 and $J^2 = J_3(J_3 + 1) + 2J_-J_+$ are Hermitian. This is because $(a_2/a_1)J_+ = a_2\bar{a}_2 + 2a_3\bar{a}_3$ is Hermitian. These nonunitary representations can be made unitary by redefining the scalar product in H. To do this it is sufficient to specify the scalar product between the basis states, and the required definition is

$$(|{}^{l'}_{m'}\rangle, |{}^{l}_{m}\rangle) = \delta_{ll'}\delta_{mm'} 2^{l-m} \frac{(l-m)!(l+m)!(l+p)!}{(2l)!}.$$

These results do not depend on the form we have taken for the metric σ; if we put $\sigma = I$, the identity, then we would have

$$z = a_3/(a_1 - ia_2).$$

This formalism also includes naturally the representations of the full orthogonal group $O(3)$, by enlarging the representation space to include axial tensors. The state of highest weight is then written

$$|{}^{l}_{l}\rangle = a_1^{l+p-1} a_{12}|0\rangle,$$

and the generators have a similar form (4.3).

Although $SU(2)$ is the covering group of $SO(3)$ it is not obvious how T_g defined by (4.2) is a representation of $g \in SU(2)$. In fact, we can recover the usual expression for $T_g, g \in SU(2)$ by substituting for a_1, a_2, a_3 with (2.3). Then $z = a_2/a_1 = \sqrt{2}(\alpha_2/\alpha_1)$ where α_1, α_2 are ordinary bosons. With the matrix

$$u = \begin{pmatrix} \alpha & \beta \\ \bar{\beta} & \bar{\alpha} \end{pmatrix}, \quad |\alpha|^2 + |\beta|^2 = 1,$$

which belongs to $SU(2)$, we identify $g \in SO(3)$ determined by $g_{11} = \alpha^2$, $g_{21} = -2\alpha\bar{\beta}$, $g_{12} = \alpha\beta$. Then from (4.2) T_u takes the form

$$T_u\phi(z) = (\alpha - \bar{\beta}z)^{2m_1}\phi\left(\frac{\beta + \bar{\alpha}z}{\alpha - \bar{\beta}z}\right),$$

which is the familiar expression for representations of $SU(2)$ in the space of polynomials ϕ of one variable z.

ACKNOWLEDGMENT

The author is grateful to Professor C.A.Hurst for guidance and frequent discussions during this research.

[1] M. A. Lohe and C. A. Hurst, J. Math. Phys. 12, 1882 (1971).
[2] V. Bargmann, Rev. Mod. Phys. 34, 829 (1962).
[3] H. Boerner, *Representations of Groups* (North-Holland, Amsterdam, 1970), Chap. VIII.
[4] M. A. Lohe and C. A. Hurst, J. Math. Phys. 14, 1955 (1973).
[5] D. P. Zhelobenko, Russ. Math. Surv. 17, 1 (1962).
[6] D. Pandres, Jr., J. Math. Phys. 6, 1098 (1965); see also D. Pandres, Jr., and D. A. Jacobson, J. Math. Phys. 9,1401 (1968).
[7] W. Miller, *Lie Theory and Special Functions* (Academic, New York, 1968), Chap. 2.
[8] K. V. Kademova and M. M. Kraev, Nucl. Phys. B 26, 2 (1971); see also K. V. Kademova and M. M. Kraev, Int. J. Theor. Phys. 6, 443 (1972); A. J. Bracken and D. A. Gray, Phys. Lett. 37B, 420 (1971).
[9] A. J. Bracken (private communication).
[10] N. Y. Vilenkin, *Special Functions and Theory of Group Representations* (A. M. S. Transl., Providence, R.I. 1968).

Phase operators and phase relations for photon states

Howard C. Volkin

National Aeronautics and Space Administration, Lewis Research Center, Cleveland, Ohio
(Received 4 May 1973; final revisions received 24 July 1973)

For a quantized mode of the radiation field, the operator whose classical analog is the ordinary phase factor of the mode amplitudes has been shown to be nonunitary. A mathematically rigorous formulation of the phase P is given on the basis of the canonical factorization theorem. All other phase quantities are defined in terms of P and $P^\dagger$. Many of the seemingly complex features of phase operators are found to be simple direct consequences of the general mathematical theory. It is easily seen that P is a partial isometry but not a unitary operator. In contrast to the amplitude operator, it is found that P is not a spectral operator and the set of phase eigenstates is not complete. Mathematically precise operator relations, including the rigorous statement of commutation rules, are developed. For each of the phase operators, a complete spectral analysis is given, with the nature of the spectrum, spectral decomposition (if any), and eigenstates shown explicitly. The values of the various phase operators are compared between phase states and coherent states.

1. INTRODUCTION

The quantum theory of coherence provides the physical foundation for the complete description of coherent radiation fields and for the analysis of devices that generate such fields. The theory has been extensively developed only in the last decade, notably by the pioneering work of Glauber[1,2] in a series of now classic papers. The need for such a theory became urgent with the advent of various masers, the first source of highly coherent electromagnetic radiation.

In classical theory, the correlation functions describing the coherence properties of the field can be expressed in terms of the probability distribution for the amplitudes of a complete set of field modes. The amplitudes are complex c-numbers, the Fourier coefficients in the modal expansion of the field, and serve as the random variables whose probability distribution describes the noise in the field. The phase and modulus of each amplitude are clearly defined quantities. In the quantum theory, amplitude operators replace the classical Fourier coefficients. Analogous field correlation functions are defined, but these functions are expressed in terms of the density operator for the quantum states of the field. Thus the coherence is no longer described by ordinary noise theory. The factorization of an amplitude into the product of a phase operator and a modulus operator is also not a direct analog of the classical case.

The idea of phase and modulus operators was used by Dirac[3] in his original paper on the quantization of the electromagnetic field. Thereafter his interpretation became the conventional one in discussing the phase of a quantized wave.[4] Following the classical correspondence closely, Dirac assumed the phase operator of a mode was a unitary operator $e^{i\Phi}$, where ϕ is the phase-angle observable. If such an observable ϕ existed, it would necessarily be conjugate to the photon number operator for the mode. The first detailed study of this question was made by Susskind and Glogower.[5] They demonstrated that the unitarity assumption leads to contradictions, which implies that a self-ajoint operator ϕ does not exist. Using heuristic methods, they showed how to circumvent this difficulty by defining self-adjoint operators C and S, which become equal in the classical limit to the sine and cosine of the phase angle. With these operators, valid uncertainty relations for phase and number can be given. Their work and more recent studies in the field are described in an excellent review article by Carruthers and Nieto,[6] to which we refer the reader for a comprehensive list of references.

Our purpose is to present a rigorous and complete mathematical development of phase operators. In Sec. 2 we review the background of physics and mathematics on which our treatment rests. The survey of relevant mathematical theory in Sec. 2B should be helpful to many, because it brings together material that is widely dispersed in the mathematical literature. The phase operator P of a quantized mode is defined in Sec. 3 on the basis of the canonical factorization theorem. The properties of $P, P^\dagger$ and the related operators C, S follow directly from the general mathematical theory. The phase operator P, which in the classical limit becomes the ordinary phase factor of the mode amplitude, is a partial isometry but clearly not a unitary operator. We use the canonical factorization theorem itself to prove that the mode amplitude is a closed operator. In the course of the mathematical formulation, we are required to determine the domain and range of each basic operator. This in turn enables us to derive precise operator relations, including the rigorous statement of commutation rules.

We continue the analysis of the operators in Sec. 4, and derive a complete description of their spectra and eigenstates. The eigenstates of P, which have not previously appeared in the literature, furnish an insight into the significance of the eigenstates of C and S, which are described by continuum eigenvectors and hence are not physically realizable states. Unlike the amplitude operator, the phase operator is not a spectral operator. For the sake of completeness, we sketch in Sec. 5 the corresponding spectral analysis for the relative phase operators. The classical analogs of these operators are the sine and cosine of the relative phase angle between a pair of modes. Finally, in Sec. 6 the values of the phase operators are compared when the mode is in a coherent state and a phase state. Similarly, the values of the relative phase are compared when both modes occupy coherent states and phase states.

In summary, our treatment of phase operators provides not only a mathematically precise formulation and extension of previous heuristic work, but shows how seemingly complex features of the phase operators follow as simple direct consequences of the general mathematical theory.

2. THEORETICAL PRELIMINARIES

A. Field-theoretic background

The gauge invariance of Maxwell's field equations means that in the case of the pure radiation field (free or noninteracting electromagnetic field) a gauge can always be chosen for the four-potential $(\Phi, \mathbf{A})$ such that the

scalar potential vanishes and the vector potential satisfies transversality, i.e., $\Phi = 0$ and

$$\nabla \cdot \mathbf{A} = 0. \tag{2.1}$$

It will be convenient here to describe the electromagnetic field by a discrete set of dynamical variables. This is accomplished by the familiar device of "box normalization," where the field is enclosed in a three-dimensional box of finite volume Ω on whose surface appropriate boundary conditions (entailing hermiticity of the Laplacian operator) are imposed. The complete orthonormal set of vector functions $\mathbf{u}_k(\mathbf{r})$ resulting from a given type of boundary condition corresponds to a set of oscillation modes for the field in Ω. When the vector potential $\mathbf{A}$ is expanded in the set of oscillation modes, the amplitudes of the different modes provide a discrete set of dynamical variables for the electromagnetic field. A well-known prescription enables us at any point to go to the limit $\Omega \to \infty$ and a continuum of variables.

In the Heisenberg picture of quantum mechanics, the vector potential operator has the same form as the classical potential function:

$$\mathbf{A}(\mathbf{r}, t) = c\sum_k (\hbar/\omega_k)^{1/2}[a_k \mathbf{u}_k(\mathbf{r})e^{-i\omega_k t} + a_k^\dagger \mathbf{u}_k^*(\mathbf{r})e^{i\omega_k t}]. \tag{2.2}$$

For a free field, the *mode amplitude operators* $a_k, a_k^\dagger$ are constant in time. In the classical limit, $\mathbf{A}(\mathbf{r}, t)$ may be regarded as an ordinary c-number function, and the pair $a_k, a_k^\dagger$ are complex conjugate numbers that remain constant in time when no charges or currents are present. The field operators $\mathbf{E} = -(1/c)\partial\mathbf{A}/\partial t$, $\mathbf{B} = \nabla \times \mathbf{A}$ satisfy Maxwell's equations for free field if $\Box^2\mathbf{A} \equiv (\nabla^2 - c^{-2}\partial_t^2)\mathbf{A} = 0$. The linear independence of different frequency terms in the expansion (2.2) implies

$$\nabla^2 \mathbf{u}_k + c^{-2}\omega_k^2 \mathbf{u}_k = 0 \tag{2.3}$$

in Ω.

Because they are solutions of the wave equation (2.3) in Ω with appropriate boundary conditions, the mode functions can be selected so as to satisfy orthonormality,

$$\int_\Omega \mathbf{u}_k^*(\mathbf{r}) \cdot \mathbf{u}_{k'}(\mathbf{r}) d^3r = \delta_{kk'}, \tag{2.4}$$

and condition (2.1) requires them to obey the transversality property,

$$\nabla \cdot \mathbf{u}_k(\mathbf{r}) = 0. \tag{2.5}$$

If the volume Ω is a cubical box of side L with periodic boundary conditions, the oscillation modes are traveling waves which we may take to be linearly polarized, in which case

$$\mathbf{u}_k(\mathbf{r}) = (L)^{-3/2}\hat{e}^{(\sigma)}e^{i\mathbf{k}\cdot\mathbf{r}},$$

where the propagation vector $\mathbf{k}$ has magnitude $|\mathbf{k}| = \omega_k/c$ according to Eq. (2.3). The unit polarization vector $\hat{e}^{(\sigma)}$ is required to be perpendicular to $\mathbf{k}$ by Eq. (2.5). Thus to a given $\mathbf{k}$ there correspond only two independent polarization directions. The triad $(\hat{e}^{(1)}, \hat{e}^{(2)}, \mathbf{k})$ is customarily taken to be right-orthogonal. We see that in the plane-wave case the index k may stand for the values of three Cartesian components of $\mathbf{k}$, the value of the index $\sigma = (1, 2)$ specifying the associated polarization direction. Of course, the discrete set of values of $\mathbf{k}$ are determined by the periodic boundary conditions.

The commutation relations for the mode amplitudes, from which follow the quantum-mechanical properties of the electromagnetic field, are those of a set of kinematically independent one-dimensional harmonic oscillators:

$$[a_k, a_{k'}^\dagger] = \delta_{kk'}1_k, [a_k, a_{k'}] = [a_k^\dagger, a_{k'}^\dagger] = 0, \tag{2.6}$$

where 1_k is equal to the identity operator on the domain of definition of the commutator, which will be defined in Sec. 3C. Using Eqs. (2.4) and (2.5) and the boundary conditions on the mode functions, we can evaluate the classical expression for the energy of the field:

$$(1/8\pi)\int_\Omega (\mathbf{E}^2 + \mathbf{B}^2)d^3r = \tfrac{1}{2}\sum_k \hbar\omega_k(a_k^\dagger a_k + a_k a_k^\dagger).$$

When the quantum conditions (2.6) are imposed, this can be written as $\sum_k \hbar\omega_k(n_k + \frac{1}{2})$, where $n_k \equiv a_k^\dagger a_k$ is the *photon number* (number of quanta) operator for the kth mode. The infinite constant term, sometimes called the *zero-point energy* or the *energy of the vacuum fluctuations*, is due to the noncommutativity of the amplitude operators and can be removed without disturbing the correspondence with the classical theory. We may take the Hamiltonian for the free field to be

$$H = \sum_k \hbar\omega_k n_k. \tag{2.7}$$

Since the amplitude operators for different modes commute, the Hamiltonian (2.7) is separable in the modes. Thus the pure radiation field is equivalent to an assembly of dynamically independent, one-dimensional harmonic oscillators, called *radiation oscillators*, and the state of each mode can be discussed independently. The eigenvalues n_k' of the operator n_k are the nonnegative integers $n_k' = 0, 1, 2, \cdots$, and the excitation energy of the kth mode $\hbar\omega_k n_k$ can have the values $0, \hbar\omega_k, 2\hbar\omega_k, \cdots$. Let the vector $|n_k'\rangle_k$ correspond to the energy eigenstate of the kth mode with energy $\hbar\omega_k n_k'$; in particular, $|0\rangle_k$ is the ground-state vector of the mode. The sequence of orthonormal vectors $\{|n_k'\rangle_k \mid n_k' = 0, 1, 2, \cdots\}$ is complete in the Hilbert space $\mathfrak{H}_k$ of all physically relizable states of the kth radiation oscillator. Hence $\mathfrak{H}_k$ is a separable Hilbert space. The states of the entire radiation field correspond to vectors in the infinite direct product space $\mathfrak{H} \equiv \otimes_k \mathfrak{H}_k$, although only a separable subspace of the nonseparable Hilbert space $\mathfrak{H}$ is the natural state space. The subspace $\mathfrak{H}^{(n)}$, $n = 0, 1, 2, \cdots$, of all n-photon states is a separable Hilbert space, and so is the infinite direct sum $\oplus_n \mathfrak{H}^{(n)}$, called the *Fock space* or *natural state space*.[7] The energy eigenstates of the entire field are the basic products

$$|\ldots, n_k', \ldots, n_l', \ldots\rangle \equiv \cdots |n_k'\rangle_k \cdots |n_l'\rangle_l \cdots.$$

It follows directly from the definition of the direct product space $\mathfrak{H}$ that $\Pi_k |\ \rangle_k$ is the zero vector of $\mathfrak{H}$ if any factor $|\ \rangle_k$ is the zero vector of its space $\mathfrak{H}_k$.

The commutation relations (2.6) imply that

$$[n_k, a_{k'}] = -\delta_{kk'}a_k, \quad [n_k, a_{k'}^\dagger] = \delta_{kk'}a_{k'}^\dagger,$$

and hence that

$$a_k|\ldots, n_k', \ldots\rangle = (n_k')^{1/2}|\ldots, n_k' - 1, \ldots\rangle, \tag{2.8a}$$

$$a_k^\dagger|\ldots, n_k', \ldots\rangle = (n_k' + 1)^{1/2}|\ldots, n_k' + 1, \ldots\rangle. \tag{2.8b}$$

Equations (2.8) display the role of $a_k^\dagger$ and a_k as photon creation and annihilation operators, respectively. Generating the excited states of the kth radiation oscillator by repeated application of the photon creation operator on the ground state, according to the prescription

$$|n_k'\rangle_k = (n_k'!)^{-1/2}(a_k^\dagger)^{n_k'}|0\rangle_k, \tag{2.9}$$

fixes the phase factor in each energy eigenvector of the kth mode relative to the ground-state vector of the mode.

In the ground or vacuum state $|0\rangle$ of the unperturbed electromagnetic field, all radiation oscillators are in their respective ground states, i.e., $|0\rangle = \Pi_k|0\rangle_k$. Alternatively, the condition

$$a_k|0\rangle = 0, \quad \text{all } k, \tag{2.10}$$

defines the state $|0\rangle$.

The coherent states[2] of the kth oscillator are the eigenstates $|\alpha\rangle_k$ of the mode amplitude a_k:

$$a_k|\alpha\rangle_k = \alpha|\alpha\rangle_k,$$

where the eigenvalue α is a complex number. The coherent states $|\ \rangle$ of the whole radiation field are the direct products of the coherent modal states, i.e., $|\ \rangle = \Pi_k|\alpha_k\rangle_k$. Expanded in energy eigenstates of the radiation oscillator, a normalized coherent modal state has the form

$$|\alpha\rangle_k = e^{-(1/2)|\alpha|^2}\sum_{n=0}^{\infty}(n!)^{-1/2}\alpha^n|n\rangle_k. \tag{2.11}$$

Note that this expression implies the phase factor in the vector $|\alpha\rangle_k$ has been fixed relative to the ground-state vector $|0\rangle_k$ by taking the scalar product $\langle 0|\alpha\rangle$ to be a real number: $\langle 0|\alpha\rangle = \exp(-\frac{1}{2}|\alpha|^2)$.

In the state $|\alpha\rangle_k$, the expectation value for finding the oscillator in the energy eigenstate $|n_k'\rangle_k$ is

$$p(n_k') = \langle\alpha|n_k'\rangle\langle n_k'|\alpha\rangle = (n_k'!)^{-1}|\alpha|^{2n'_k}e^{-|\alpha|^2}, \tag{2.12}$$

which is a Poisson distribution with mean value

$$\langle n_k\rangle \equiv \langle\alpha|n_k|\alpha\rangle = \sum n_k' p(n_k') = |\alpha|^2. \tag{2.13}$$

The second moment of the Poisson distribution is

$$\langle n_k^2\rangle \equiv \langle\alpha|n_k^2|\alpha\rangle = \sum n_k'^2 p(n_k') = |\alpha|^4 + |\alpha|^2,$$

so that the uncertainty in the value of n_k is

$$\Delta n_k = (\langle n_k^2\rangle - \langle n_k\rangle^2)^{1/2} = |\alpha| = \langle n_k\rangle^{1/2}. \tag{2.14}$$

In general, for the classical theory to describe a mode of the radiation field, it is necessary that the state of the mode be a superposition of many n-photon states involving values of $n \gg 1$. As Eqs. (2.13) and (2.14) suggest, the coherent states describable in the classical domain are those with values of α in the region $|\alpha| \gg 1$ of the complex α plane.

B. Mathematical background

A linear operator V in a Hilbert space $\mathfrak{H}$ is a linear mapping of a subspace $\mathfrak{D}(V) \equiv \mathfrak{M}$, the domain of definition of V, onto a subspace $\mathfrak{R}(V) \equiv \mathfrak{N}$, the range of values of V. A linear operator V is called an *isometric* operator or an *isometry* if $\langle V\psi|V\psi\rangle = \langle\psi|\psi\rangle$ for all $\psi \in \mathfrak{M}$ or, equivalently, if $\langle V\psi_1|V\psi_2\rangle = \langle\psi_1|\psi_2\rangle$ for any pair $\psi_1, \psi_2 \in \mathfrak{M}$. Since $V\psi = 0$ implies $\psi = 0$, an isometry V is necessarily a one-to-one mapping of $\mathfrak{M}$ onto $\mathfrak{N}$ and possesses an inverse V^{-1}, the reverse one-to-one isometric mapping of $\mathfrak{N}$ onto $\mathfrak{M}$. If the domain $\mathfrak{M}$ of an isometry V is dense in $\mathfrak{H}$, then the adjoint operator $V^\dagger$ is defined and the domain of $V^\dagger$ contains at least the range $\mathfrak{N}$ of V, where it is equal to V^{-1}. Thus $V^\dagger$ maps $\mathfrak{N}$ isometrically onto $\mathfrak{M}$. On $\mathfrak{M}$ the operator $V^\dagger V$ is equal to the identity operator, and on $\mathfrak{N}$ the operator $VV^\dagger$ is equal to the identity operator. Clearly, an isometry V is bounded on its domain $\mathfrak{M}$ with norm $|V| = 1$.

An isometry V is said to be *maximal* if either its domain $\mathfrak{M}$ or its range $\mathfrak{N}$ is equal to the whole Hilbert space $\mathfrak{H}$. A maximal isometry clearly has no proper isometric extension in $\mathfrak{H}$. Let I denote the identity operator on $\mathfrak{H}$. If $\mathfrak{M} = \mathfrak{H}$, then $V^\dagger V = I$; if $\mathfrak{N} = \mathfrak{H}$, then $VV^\dagger = I$. (The equality sign between operators implies that their domains are equal.) If U is an isometry such that both its domain and range are equal to $\mathfrak{H}$, i.e., if U is an isometric mapping of $\mathfrak{H}$ onto itself, then U is a *unitary* operator. For a unitary operator, $U^\dagger$ is also unitary and $U^\dagger U = UU^\dagger = I$.

A *partial isometry* P is a bounded linear operator on the entire Hilbert space $\mathfrak{H}$ with the property that there exists a closed subspace $\mathfrak{M}$ of $\mathfrak{H}$ such that

(1) P is an isometry on $\mathfrak{M}$, i.e., $\langle P\psi|P\psi\rangle = \langle\psi|\psi\rangle$ for all $\psi \in \mathfrak{M}$, and

(2) the orthogonal complement $\mathfrak{M}^\perp$ is the null space of P, i.e., P maps $\mathfrak{M}^\perp$ onto the element zero: $P\mathfrak{M}^\perp = \{0\}$.

Since the subspace $\mathfrak{M}$ is a closed set, its isometric image $\mathfrak{N} = \mathfrak{R}(P) = P\mathfrak{H} = P\mathfrak{M}$ is necessarily a closed subspace also. The subspace $\mathfrak{M}$ is called the *initial domain* of P and the range $\mathfrak{N}$ of P is called the *final domain*. The restriction P' of P to $\mathfrak{M}$ is an isometry, so that for P': $\mathfrak{M} \to \mathfrak{N}$, the inverse exists, $(P')^{-1}$: $\mathfrak{N} \to \mathfrak{M}$. Let Q be the operator that is equal to $(P')^{-1}$ on $\mathfrak{N}$ and equal to the zero operator on $\mathfrak{N}^\perp$. Thus Q is a partial isometry, with initial domain $\mathfrak{N}$ and final domain $\mathfrak{M}$.

The following properties of a partial isometry P are direct consequences of its definition[8]:

(i) P is bounded on $\mathfrak{H}$ with norm $|P| = 1$;

(ii) $P^\dagger = Q$, $P = Q^\dagger$;

(iii) $P^\dagger P$ is the orthogonal (self-adjoint) projection onto the subspace $\mathfrak{M}$, i.e.,

$$P^\dagger P = I - \mathcal{P}, \quad (P^\dagger P)^2 = P^\dagger P, \quad \mathfrak{R}(P^\dagger P) = (P^\dagger P)\mathfrak{H} = \mathfrak{M},$$

where $\mathcal{P}$ is the orthogonal projection onto $\mathfrak{M}^\perp$;

(iv) $PP^\dagger$ is the orthogonal projection onto $\mathfrak{N}$, i.e.,

$$PP^\dagger = I - \mathcal{Q}, \quad (PP^\dagger)^2 = PP^\dagger, \quad \mathfrak{R}(PP^\dagger) = \mathfrak{N},$$

where $\mathcal{Q}$ is the projection onto $\mathfrak{N}^\perp$;

(v) a necessary and sufficient condition that a bounded linear operator P be a partial isometry is that $P^\dagger P$ be a projection;

(vi) the operator P is a partial isometry if and only if $P^\dagger$ is a partial isometry.

Our development of phase operators will be based on the canonical factorization (polar decomposition) theorem[8,9]. A closed transformation T with domain $\mathfrak{D}(T)$ dense in the Hilbert space $\mathfrak{H}$ can be written in one and only one way as a product $T = PA$, where P is a partial isometry whose initial domain is $\overline{\mathfrak{R}(T^\dagger)}$, the closure of the range of $T^\dagger$, and A is a positive self-adjoint operator such that $\overline{\mathfrak{R}(A)} = \overline{\mathfrak{R}(T^\dagger)}$. The operator A is the positive square root of the positive self-adjoint operator $T^\dagger T$: $A = (T^\dagger T)^{1/2}$.

The equality $T = PA$ then implies $\mathfrak{D}(A) = \mathfrak{D}(T)$ and $\mathfrak{R}(P) = \overline{\mathfrak{R}(T)}$. If the operator T in the theorem is a normal operator, i.e., if $T^\dagger T = TT^\dagger$, then $T = PA = AP$ and P is unitary.

We now briefly review those properties of unbounded operators which bear most directly on the theorem and our application of it. A linear operator T in a Hilbert space $\mathfrak{H}$ can have an adjoint only if its domain is dense in $\mathfrak{H}$. Thus whenever we write $T^\dagger$, it is to be understood that $\mathfrak{D}(T)$ is dense. If B is a bounded operator on the entire space $\mathfrak{H}$, then $\mathfrak{D}(B+T)=\mathfrak{D}(BT)=\mathfrak{D}(T)$, $(B+T)^\dagger = B^\dagger + T^\dagger$, and $(BT)^\dagger = T^\dagger B^\dagger$. A self-adjoint operator $A = A^\dagger$ is maximal symmetric, i.e., A has no symmetric (and hence no self-adjoint) extensions in $\mathfrak{H}$. If T is a positive self-adjoint operator, there is a unique positive self-adjoint operator A such that $A^2 = T$; we call A the *positive square root of* T and write $A = T^{1/2}$. The adjoint $T^\dagger$ is a closed operator; in particular, a self-adjoint operator $A = A^\dagger$ is necessarily closed.

If T is a linear operator with dense domain in $\mathfrak{H}$, then $\mathfrak{D}(T^\dagger)$ is dense in $\mathfrak{H}$ if and only if T possesses a closed linear extension; then $T^{\dagger\dagger} \equiv (T^\dagger)^\dagger$ is the minimal closed linear extension of T, i.e., every closed linear extension of T is also an extension of $T^{\dagger\dagger}$. The extensions need not be proper ones, of course. In particular, if the operator T itself is closed, then $T^{\dagger\dagger} = T$. Moreover, if T is closed and has dense domain, $T^\dagger T$ is self-adjoint and positive and $(I + T^\dagger T)^{-1}$ exists, is everywhere defined and bounded on $\mathfrak{H}$, and is self-adjoint. If T is closed and unbounded, its domain cannot be the entire Hilbert space, i.e., $\mathfrak{D}(T)$ must be a proper subspace of $\mathfrak{H}$. This follows from the closed-graph theorem, which tells us that a closed operator that is everywhere defined in $\mathfrak{H}$ must be bounded.

The unbounded operators that occur here will be found to be closed operators. In this connection, we note that a bounded operator B is closed if and only if its domain $\mathfrak{D}(B)$ is a closed set. In particular, B is closed whenever $\mathfrak{D}(B) = \mathfrak{H}$. In order to discuss the spectrum of a closed operator T, we consider the operator $T - \lambda I$, where λ is a complex number. The operator $T - \lambda I$ has domain $\mathfrak{D}(T)$ and is closed if T is closed. The *resolvent set* $\rho(T)$ consists of those values of λ for which $T - \lambda I$ is a one-to-one mapping of $\mathfrak{D}(T)$ onto all of $\mathfrak{H}$. Now, whenever the inverse of a closed operator exists, the inverse is also closed. Hence, $\rho(T)$ is the set of complex numbers such that $(T - \lambda I)^{-1}$ exists and is a bounded linear operator defined everywhere in $\mathfrak{H}$. *The spectrum* $\sigma(T)$ of T is defined as the complement of $\rho(T)$ in the complex plane. It can be shown that the spectrum of a closed operator is a closed set. Thus a closed operator T partitions the complex plane into the set $\rho(T)$ and the closed set $\sigma(T)$.

The spectrum $\sigma(T)$ consists of three nonoverlapping sets called the *point spectrum*, the *continuous spectrum*, and the *residual spectrum* and denoted by $\sigma_p(T)$, $\sigma_c(T)$, and $\sigma_r(T)$, respectively. These sets are determined by the nature of the mapping $(T - \lambda I)$ on its domain $\mathfrak{D}(T)$ and of the range subspace $\mathfrak{R}(T - \lambda I)$ onto which $\mathfrak{D}(T)$ is mapped. The point spectrum is the set of all complex numbers λ in $\sigma(T)$ for which $(T - \lambda I)$ is not a one-to-one mapping of $\mathfrak{D}(T)$. Thus, $\lambda \in \sigma_p(T)$ if and only if there exists at least one vector $\psi \neq 0$ in $\mathfrak{H}$ such that $(T - \lambda I)\psi = 0$, or equivalently, if and only if T has at least one eigenvector in $\mathfrak{H}$ with the eigenvalue λ. If λ belongs to either $\sigma_c(T)$ or $\sigma_r(T)$, $(T - \lambda I)$ is a one-to-one mapping of $\mathfrak{D}(T)$ onto $\mathfrak{R}(T - \lambda I)$, but the range is not equal to $\mathfrak{H}$. If $\lambda \in \sigma_c(T)$, the subspace $\mathfrak{R}(T - \lambda I)$ is dense in but not equal to $\mathfrak{H}$; if $\lambda \in \sigma_r(T)$, the subspace $\mathfrak{R}(T - \lambda I)$ is not dense in $\mathfrak{H}$.

If T is a bounded operator on $\mathfrak{H}$, its spectrum $\sigma(T)$ is a closed, nonvoid subset of the circle $|\lambda| \leq |T|$. If T is unbounded, its spectrum may be a bounded set, an unbounded set, the void set, or even the whole complex plane. The spectrum of a unitary operator, a self-adjoint operator or a positive self-adjoint operator lies on the unit circle, the real axis, or the nonnegative real axis, respectively. A self-adjoint operator A is bounded if and only if its spectrum is a bounded set; if A is bounded, its spectrum is a bounded closed set on the real axis. A unitary or a self-adjoint (indeed, any normal) operator has no residual spectrum. If the Hilbert space $\mathfrak{H}$ is separable, the point spectrum of a unitary or a self-adjoint operator is discrete (i.e., finite or countably infinite).

To describe completely a spectral operator T and a resolution of the identity for T requires a rather lengthy series of technical definitions. We present here only a bare outline with emphasis on some important features that will be encountered later. A *spectral measure* $E(\delta)$, $\delta \in \mathfrak{B}$, in $\mathfrak{H}$ is a mapping of a Boolean algebra of sets $\mathfrak{B}$ into a Boolean algebra of projection operators in $\mathfrak{H}$ such that $E(1) = I$, where 1 is the unit set in $\mathfrak{B}$. A *resolution of the identity (spectral decomposition)* for the closed operator T with dense domain in $\mathfrak{H}$ is a countably additive spectral measure E defined on the Borel sets of the plane and related to T in a way indicated by the symbolic representation

$$T = \int_{\sigma(T)} \lambda E(d^2\lambda), \tag{2.15}$$

where $d^2\lambda$ is the (real) differential element of area in the complex λ plane. We note that $E(\delta) = I$ if $\delta \supseteq \sigma(T)$, and $E(\delta) = 0$ if the intersection $\delta \cap \sigma(T)$ is the void set. A closed and densely defined operator that possesses a resolution of the identity is called a *spectral operator*. The resolution of the identity for a spectral operator is unique.

A spectral operator for which Eq. (2.15) is a well-defined operator equation (rather than a symbolic representation) is called a *scalar type* spectral operator.[10] Associated with the resolution of the identity E for a scalar type spectral operator T is a functional calculus of operators. If f is a complex valued Borel measurable function of a complex variable, then

$$f(T) = \int_{\sigma(T)} f(\lambda) E(d^2\lambda) \tag{2.16}$$

is a closed linear operator with domain defined by

$$(f(T)\psi, f(T)\psi) = \int_{\sigma(T)} |f(\lambda)|^2 (\psi, E(d^2\lambda)\psi) < \infty,$$

$$\psi \in \mathfrak{D}(f(T)).$$

The amplitude operator a_k is an example of a scalar type spectral operator. Glauber[2] has shown that the coherent states $\{|\alpha\rangle_k\}$ of the kth mode oscillator form a complete set in $\mathfrak{H}_k$ by deriving the completeness relation as a certain integral over the complex α plane of the projection operators $|\alpha\rangle\langle\alpha|$:

$$(1/\pi)\int |\alpha\rangle_{k\ k}\langle\alpha| d^2\alpha = I_k. \tag{2.17}$$

This takes a familiar form even though the set of coherent states is not orthogonal. According to Eq. (2.17), any vector in $\mathfrak{H}_k$ can be expressed as a linear combination of $|\alpha\rangle_k$, and hence the set of coherent states is complete. But $\mathfrak{H}_k$ has a denumerable orthogonal basis (see Sec. 3B). Therefore, the set of coherent states, which is uncountably infinite, cannot be linearly independent or orthogonal. It follows that the spectral decomposition of a_k is

$$a_k = (1/\pi)\int \alpha|\alpha\rangle_{k\ k}\langle\alpha| d^2\alpha. \tag{2.18}$$

Thus a_k has a pure point spectrum which consists of every (finite) point in the complex plane.

A familiar spectral-operator in quantum mechanics, of course, is the self-adjoint operator A representing a physical quantity, and the properties of the resolution of the identity for A are well known. The spectral measure belonging to a unitary operator vanishes on any Borel set that contains no points of the unit circle (on which the spectrum lies). The spectral theorem for unitary operators[11] tells us that any unitary operator U has a resolution of the identity of the form

$$U = \int_0^{2\pi} e^{i\theta} dE(\theta), \tag{2.19}$$

where the family of projections $E(\theta)$ varies only in the interval $0 \leq \theta \leq 2\pi$ and, as a function of the real variable θ, possesses all the properties required of a family of projections that belongs to a self-adjoint operator. The requirement that $E(\theta)$ be continuous at $\theta = 0$, i.e., $E(\theta) = 0$, $\theta \leq 0$, determines $E(\theta)$ uniquely. The self-adjoint operator defined by $A = \int_0^{2\pi} \theta dE(\theta)$ is bounded on $\mathfrak{H}$ and its spectrum is contained in the interval $[0, 2\pi]$. By Eq. (2.16) the operator U is the exponential function of A: $U = e^{iA}$. An isometric operator that is maximal but not unitary has no resolution of the identity and is not a spectral operator.[12]

For a comprehensive treatment of the mathematical theory outlined in this section, we refer the reader to the volumes of Dunford and Schwartz[13] and to Riesz and Sz.-Nagy.[14] To complete our survey of the mathematical theory used here, we mention the alternative treatment of the continuous spectrum, the well-known method of Dirac, whose rigorous justification is given by the theory of distributions. In Dirac's method, the continuous spectrum of a physical quantity (self-adjoint operator) is treated formally like the point spectrum. However it is necessary to employ a vector space $\mathfrak{V}$ with a generalized scalar product, which in some cases is a singular distribution, namely, the Dirac delta function. The Hilbert space of physical states $\mathfrak{H}$ (which contains the eigenvectors associated with the eigenvalues of the point spectrum) is embedded in $\mathfrak{V}$, but $\mathfrak{V}$ is not a Hilbert space because of the singular nature of its scalar product. The eigenvectors associated with an eigenvalue in the continuum belong to $\mathfrak{V}$, but such vectors do not represent physically realizable states and are not contained entirely in $\mathfrak{H}$. It is in its dependence on two continuum eigenvectors that the scalar product is a singular function. The completeness relation for a self-adjoint operator A in the Dirac formalism is

$$\sum_{a'} |a'\rangle\langle a'| + \int da'' |a''\rangle\langle a''| = \mathcal{I},$$

where $a' \in \sigma_p(A)$, $a'' \in \sigma_c(A)$, and $\mathcal{I}$ is the identity operator in $\mathfrak{V}$. The restriction of $\mathcal{I}$ to the subspace $\mathfrak{H}$ is equal to I, of course.

3. PHASE OPERATORS

A. The modal phase operator

The amplitudes a_k and $a_k^\dagger$ are both closed operators defined on a dense domain in $\mathfrak{H}_k$ (as will be shown in Sec. 3B). Hence both operators satisfy the conditions of the canonical factorization theorem. Applying the theorem to a_k, we obtain the unique factorizations

$$a_k = P_k n_k^{1/2}, \quad a_k^\dagger = n_k^{1/2} P_k^\dagger \tag{3.1}$$

for a_k and its adjoint. The self-adjoint factor must be the positive square root of the photon number operator for the kth mode: $n_k = a_k^\dagger a_k$. In view of the Eqs. (2.8a, b) satisfied by the amplitude operators, it is apparent that the partial isometry P_k and its adjoint, the partial isometry $P_k^\dagger$, are defined on $\mathfrak{H}_k$ by

$$P_k |n_k'\rangle_k = |n_k' - 1\rangle_k, \quad n_k' \geq 1, \tag{3.2a}$$

$$P_k |0\rangle_k = 0, \tag{3.2b}$$

$$P_k^\dagger |n_k'\rangle_k = |n_k' + 1\rangle_k. \tag{3.2c}$$

We call P_k and $n_k^{1/2}$ the *phase operator* and *modulus operator*, respectively, of the kth mode. Note that P_k and $P_k^\dagger$ act as "discrete displacement" operators (lowering and raising operators, respectively) on the energy eigenstates. The phase operator P_k is bounded with norm $|P_k| = 1$. Therefore its spectrum is contained in the closed unit circle: $\sigma(P_k) \subseteq C$, where $C = \{u \,||u| \leq 1\}$. We shall see that the point spectrum of P_k is the interior of the unit circle and the continuous spectrum is the boundary of the unit circle, i.e., $\sigma_p(P_k) = \{u \,||u| < 1\}$, $\sigma_c(P_k) = \{u \,||u| = 1\}$. The adjoint $P_k^\dagger$ has norm $|P_k^\dagger| = |P_k| = 1$. However, the spectrum of $P_k^\dagger$ is a purely residual one, i.e., $\sigma_r(P_k^\dagger) = \{u \,||u| \leq 1\}$. The eigenstates of n_k are also eigenstates of $n_k^{1/2}$. Since these eigenstates form a complete set in $\mathfrak{H}_k$, the spectrum of the self-adjoint operator $n_k^{1/2}$, like that of n_k, is a pure point spectrum; it consists of the set of nonnegative numbers $\sigma(n_k^{1/2}) = \{n'^{1/2} \,|n' = 0, 1, 2, \cdots\}$. When the classical description of the mode applies, the values of $n_k^{1/2}$ will be large positive numbers and the values of P_k will be found in the region $|u| \lesssim 1$. Therefore, in the classical limit, where a_k can be treated as an ordinary complex number, the operators $n_k^{1/2}$ and P_k do indeed play the roles of modulus and phase factor, respectively, of a_k.

The initial domain of the partial isometry P_k is $\mathfrak{M}_k = \overline{\mathcal{R}(a_k^\dagger)} = \overline{\text{sp}}\{|n_k'\rangle_k \,|n_k' \geq 1\}$, the closure of the subspace $\mathcal{R}(a_k^\dagger)$ spanned by the set of energy eigenstates for which $n_k' \neq 0$ in agreement with Eq. (3.2a). The one-dimensional subspace or ray spanned by the ground state of the kth radiation oscillator is the orthogonal complement of $\mathfrak{M}_k$ in $\mathfrak{H}_k$: $\mathfrak{M}_k^\perp = \text{sp}\{|0\rangle_k\}$. Hence $\mathfrak{M}_k^\perp$ is the null space of P_k, which accords with Eq. (3.2b). The final domain of P_k, which is equal to the initial domain of the partial isometry $P_k^\dagger$, is the entire Hilbert space, i.e., $\mathfrak{N}_k = \mathfrak{H}_k$. From properties (iii) and (iv) of Sec. 2B, we have

$$P_k^\dagger P_k = I_k - \mathcal{P}_k, \quad P_k P_k^\dagger = I_k, \tag{3.3}$$

$$[P_k, P_k^\dagger] = \mathcal{P}_k \equiv |0\rangle_k \,{}_k\langle 0|, \tag{3.4}$$

where $\mathcal{P}_k$ is the projection operator onto the ray corresponding to the ground state.

Let us now apply the canonical factorization theorem to the adjoint $a_k^\dagger$. This case requires the positive square root of the self-adjoint operator $a_k a_k^\dagger = n_k + 1_k$. In the next section, we will show that the domains of $a_k a_k^\dagger$ and $n_k = a_k^\dagger a_k$ are equal, which means that they are equal to the domain of the commutator (2.6). The definition of the sum operator then permits us to write $a_k a_k^\dagger = n_k + I_k$. The theorem then gives the unique expressions

$$a_k^\dagger = P_k^\dagger (n_k + I_k)^{1/2}, \; a_k = (n_k + I_k)^{1/2} P_k, \tag{3.5}$$

where we have used $(a_k^\dagger)^\dagger = a_k$, which is valid for any closed and densely defined operator.

As an isometry, P_k has range $\mathcal{R}(P_k) = \mathfrak{H}_k$ and domain $\mathcal{D}(P_k) = \mathfrak{M}_k \neq \mathfrak{H}_k$. Hence P_k is maximal isometric but not unitary. The fact that its range is equal to all of $\mathfrak{H}_k$ precludes an isometric extension of P_k to all of $\mathfrak{H}_k$

and ultimately rules out the possibility of P_k having a unitary extension. The commutator (3.4) is an expression of this; the existence of the ground state requires that the partial isometry P_k possess a nonzero null space and not be unitary. Obviously, a is not a normal operator, as the commutation relations (2.6) explicitly state, and hence P is not required to be unitary. Moreover, the existence of a ground state $|0\rangle$ satisfying Eq. (2.10) is due basically to the commutation relations (2.6), from which Eq. (2.8a) derives. Thus the nonunitary nature of P stems ultimately from the commutation relations (2.6).

Since it is bounded on $\mathfrak{H}_k$, P_k is a closed operator and therefore satisfies the conditions of the canonical factorization theorem. Thus we can write $P_k = V_k(P_k^\dagger P_k)^{1/2}$. Under circumstances where the ground state can be ignored and only operator restrictions to $\mathfrak{M}_k$ enter, we can regard the operators P_k and $P_k^\dagger$ as commutative. Then P_k and V_k can be regarded as equal and unitary, and the values of P_k lie essentially on the unit circle. This situation occurs whenever the classical description applies.

B. Properties of the operators

In this section, we will retrace our course somewhat in order to verify the basic properties of our operators that have already been utilized. This not only will justify the results obtained but will illuminate the entire mathematical structure. In discussing a single mode of the radiation field, we may simplify the notation by dropping the mode index k as a subscript. Henceforth we will feel free to omit the mode subscript k when only a single mode is involved.

As is well known, the self-adjoint operator $n = a^\dagger a$ in the Hilbert space $\mathfrak{H}$ has a pure point spectrum, which consists of the nonnegative integers. Also, the spectrum is simple, i.e., the eigenvalues are all of multiplicity 1 (nondegenerate). If follows from the spectral theorem for self-adjoint operators that the set of orthonormal energy eigenstates $\{|n'\rangle | n' = 0, 1, 2, \cdots\}$ is complete in $\mathfrak{H}$, and n has the spectral decomposition

$$n = \sum_{n'=0}^{\infty} n' |n'\rangle\langle n'|, \qquad \sum_{n'} |n'\rangle\langle n'| = I.$$

The entire Hilbert space is the closure of the subspace spanned by the energy eigenstates and can be defined as the set

$$\mathfrak{H} = \left\{ \psi = \sum_{n'=0}^{\infty} c_{n'} |n'\rangle \;\middle|\; \sum_{n'=0}^{\infty} |c_{n'}|^2 < \infty \right\}.$$

The partial isometries P and $P^\dagger$ are bounded operators on $\mathfrak{H}$. The initial domain of P in Eq. (3.1) is the set

$$\mathfrak{M} = \left\{ \psi = \sum_{n'=1}^{\infty} c_{n'} |n'\rangle \;\middle|\; \sum_{n'=1}^{\infty} |c_{n'}|^2 < \infty \right\}.$$

The domain of n is defined as the set of all vectors ψ such that

$$|n\psi|^2 = \sum_{n'=0}^{\infty} n'^2 \langle\psi|n'\rangle\langle n'|\psi\rangle = \sum_{n'=0}^{\infty} n'^2 |c_{n'}|^2 < \infty, \qquad \psi \in \mathfrak{D}(n). \quad (3.6)$$

In the operational calculus based on the resolution of the identity belonging to n, the positive square root of n is expressed as $n^{1/2} = \sum_{n'=0}^{\infty} n'^{1/2} |n'\rangle\langle n'|$ and its domain is defined by

$$|n^{1/2}\psi|^2 = \sum_{n'=0}^{\infty} n' \langle\psi|n'\rangle\langle n'|\psi\rangle = \sum_{n'=0}^{\infty} n' |c_{n'}|^2 < \infty, \qquad \psi \in \mathfrak{D}(n^{1/2}). \quad (3.7)$$

Note that $\mathfrak{D}(n)$ is a proper subspace of $\mathfrak{D}(n^{1/2})$. Clearly, the domain of the operator $aa^\dagger$ is equal to that of n: $\mathfrak{D}(aa^\dagger) = \mathfrak{D}(n)$.

We now consider the amplitude operators. Let us at first regard Eqs. (2.8a, b) as defining relations for a pair of operators a and $a^\dagger$ in $\mathfrak{H}$ without assuming the adjoint relation between them. Both operators are clearly unbounded (discontinuous), and we must specify their domains of definition. By linearity, the relations (2.8) define both operators on the subspace $\mathfrak{D}$ spanned by the set of all energy eigenstates of the oscillator: $\mathfrak{D} \equiv \mathrm{sp}\{|n'\rangle | n' = 0, 1, 2, \cdots\}$. Since the set is complete, the closure of $\mathfrak{D}$ is the entire Hilbert space $\mathfrak{H}$ and $\mathfrak{D}$ is dense in $\mathfrak{H}$. The domain of definition of a and $a^\dagger$ can be extended to the subspace $\mathfrak{D}(a) = \mathfrak{D}(n^{1/2})$ defined by condition (3.7). If ψ is an infinite linear combination, $\psi = \sum_{n'} c_{n'} |n'\rangle$, we define the operators by

$$a\psi = \sum_{n'} c_{n'} (n')^{1/2} |n' - 1\rangle,$$

$$a^\dagger\psi = \sum_{n'} c_{n'} (n' + 1)^{1/2} |n' + 1\rangle.$$

If $\varphi = \sum_{n'} b_{n'} |n'\rangle$ belongs to $\mathfrak{D}(a)$, then

$$\langle\varphi|a\psi\rangle = \langle a^\dagger\varphi|\psi\rangle = \sum_{n'=1}^{\infty} (n')^{1/2} b_{n'-1}^* c_{n'}$$

holds for any pair $\varphi, \psi \in \mathfrak{D}(a)$, since the series is convergent by the Cauchy–Schwarz inequality. It follows immediately from the definition of the adjoint operator that $a^\dagger$ is equal to the adjoint of a on $\mathfrak{D}(a)$ and vice-versa. Since the common domain $\mathfrak{D}(a)$ of a and $a^\dagger$ is dense, a necessarily possesses a closed linear extension (although not necessarily a proper closed extension) and $(a^\dagger)^\dagger \equiv a^{\dagger\dagger}$ is the minimal closed linear extension of a, i.e., $a^{\dagger\dagger}$ is equal to a on $\mathfrak{D}(a)$ and every closed linear extension of a is also an extension of $a^{\dagger\dagger}$. The closed operator $a^{\dagger\dagger}$ has a dense domain, since it is defined at least on the dense set $\mathfrak{D}(a)$, where it is equal to a. Thus $a^{\dagger\dagger}$ satisfies the conditions of the canonical factorization theorem.

Now the operation of a is given everywhere on $\mathfrak{D}(a)$ by the expression (3.1) when the operation of P on the energy eigenstates is defined by (3.2a, b). Since $\mathfrak{D}(P) = \mathfrak{H}$, the domain of the product $Pn^{1/2}$ is $\mathfrak{D}(n^{1/2}) = \mathfrak{D}(a)$. Thus the form (3.1) is a valid representation of a everywhere on $\mathfrak{D}(a)$. The uniqueness of the canonical factorization means that the factorization of $a^{\dagger\dagger}$ takes the same form (3.1) on $\mathfrak{D}(a)$. But a self-adjoint operator such as $n^{1/2}$ has no self-adjoint extension. Hence the form (3.1) cannot be extended to a larger domain, and the domain of $a^{\dagger\dagger}$ is equal to $\mathfrak{D}(a)$. Therefore, $a^{\dagger\dagger} = a$, the operator a is closed, and the pair of operators $a, a^\dagger$ are the adjoints of one another.

Definitions (3.6) and (3.7) indicate that both $\mathfrak{D}(n)$ and $\mathfrak{D}(n^{1/2})$ are invariant under the operation of $P^\dagger$. Hence the domain of $n^{1/2}P^\dagger$ is clearly equal to $\mathfrak{D}(n^{1/2}) = \mathfrak{D}(a) = \mathfrak{D}(a^\dagger)$, which completes the verification of the second operator equality in (3.1).

Not only the domains of $n^{1/2}$ and a but also their ranges are equal. To see this, observe that

$$\psi = \sum_{n'=0}^{\infty} c_{n'} |n'\rangle \qquad \varphi = \sum_{n'=0}^{\infty} \left(\frac{n'}{n'+1}\right)^{1/2} c_{n'} |n' + 1\rangle,$$

are either both contained in $\mathfrak{D}(n^{1/2}) = \mathfrak{D}(a)$ or neither is. Since $n^{1/2}\psi = a\varphi$, it follows that $\mathfrak{R}(n^{1/2}) = \mathfrak{R}(a)$. Unlike the operator n, which has the non-zero null space $\mathfrak{N}^{\perp}$, the operator $n + I$ is a one-to-one mapping of its domain $\mathfrak{D}(n)$ onto its range. Likewise the operator $(n + I)^{1/2}$ is a one-to-one mapping of its domain $\mathfrak{D}(n^{1/2}) = \mathfrak{D}(a)$ onto $\mathfrak{R}((n + I)^{1/2}) = \mathfrak{R}(a)$. The inverse operator provides alternative forms for suitable restrictions of the operators P and $P^{\dagger}$:

$$P\psi = (n + I)^{-1/2}a\psi, \quad \text{all } \psi \in \mathfrak{D}(a),$$

$$P^{\dagger}\varphi = a^{\dagger}(n + I)^{-1/2}\varphi, \quad \text{all } \varphi \in \mathfrak{R}(a).$$

C. Operator relations

The operator $aa^{\dagger}$ has domain of definition $\mathfrak{D}(n)$, and hence the commutator $[a, a^{\dagger}]$ has also the domain $\mathfrak{D}(n)$. Thus the operator 1 that appears in the commutation relation (2.6) stands for the restriction to $\mathfrak{D}(n)$ of the identity I on $\mathfrak{H}$ and makes the domains of both sides of the equation equal. We can rewrite Eq. (2.6) in the form

$$aa^{\dagger} = n + I, \tag{3.8}$$

since both operators of the equality have the same domain.

Now the equation $n = n^{1/2}n^{1/2}$ implies that $\mathfrak{D}(n)$ is the set of all vectors in $\mathfrak{H}$ that are mapped by $n^{1/2}$ into $\mathfrak{D}(n^{1/2})$. Furthermore, the subspace $\mathfrak{D}(n)$ is invariant under $P^{\dagger}$, and $P^{\dagger}$ maps no other vectors of $\mathfrak{H}$ into $\mathfrak{D}(n)$. Accordingly, the forms (3.1) for a and $a^{\dagger}$ can be inserted into Eq. (3.8) and give

$$PnP^{\dagger} = n + I, \tag{3.9}$$

More generally, in the operational calculus based on the resolution of the identity belonging to n, we have

$$Pf(n)P^{\dagger} = \sum_{n'=0}^{\infty} f(n')P|n'\rangle\langle n'|P^{\dagger} = \sum_{n'=1}^{\infty} f(n')|n' - 1\rangle\langle n' - 1|$$

$$= \sum_{n'=0}^{\infty} f(n' + 1)|n'\rangle\langle n'|, \tag{3.10}$$

$$Pf(n)P^{\dagger} = f(n + I).$$

Equation (3.9) reveals again the nonunitary character of the partial isometry P. If P were unitary, the transformation of n on the left of (3.9) would be a similarity transformation and hence would preserve the spectrum of n, which is clearly not the case.

A vector ψ in $\mathfrak{H}$ is mapped by P into $\mathfrak{D}(n)$ if and only if ψ belongs to $\mathfrak{D}(n)$. Therefore, multiplying Eq. (3.9) on the right by P is permissible and retains $\mathfrak{D}(n)$ as the domain of the resulting equation: $PnP^{\dagger}P = nP + P$. Using (3.3) and the operator equation $n\mathcal{P} = 0$, we obtain

$$Pn = nP + P, \quad [P, n] = P1, \tag{3.11}$$

where again the operator 1 is used in the equivalent commutator form of the equation in order to indicate the restriction to $\mathfrak{D}(n)$. Similarly, multiplying Eq. (3.9) on the left by $P^{\dagger}$ and using relations (3.3) together with the fact that on $\mathfrak{D}(n)$ the operator $\mathcal{P}n$ is equal to the zero operator gives

$$[P^{\dagger}, n] = -P^{\dagger}1, \tag{3.12}$$

the adjoint of Eq. (3.10).

It is evident that the subspace $\mathfrak{D}(n^{1/2}) = \mathfrak{D}(a) = \mathfrak{D}(a^{\dagger})$ is invariant under either P or $P^{\dagger}$, while no other vectors of $\mathfrak{H}$ are mapped into $\mathfrak{D}(n^{1/2})$ by either operator. The equations

$$Pn^{1/2} = (n + I)^{1/2}P, \tag{3.13a}$$

$$n^{1/2}P^{\dagger} = P^{\dagger}(n + I)^{1/2}, \tag{3.13b}$$

obtained by equating the two forms of a and $a^{\dagger}$ given by relations (3.1) and (3.5), both have the domain $\mathfrak{D}(n^{1/2})$. We can safely multiply Eq. (3.13a) on the right by $P^{\dagger}$ and obtain, with the help of (3.3),

$$Pn^{1/2}P^{\dagger} = (n + I)^{1/2},$$

which is a special case of Eq. (3.10).

The partial isometry P is definitely not unitary. It is instructive, however, to trace the main consequences that would derive were P a unitary operator. From the spectral theorem for unitary operators, it would follow that P is expressible in the form $e^{i\phi}$, where ϕ is a bounded self-adjoint operator on $\mathfrak{H}$ with spectrum in the interval $[0, 2\pi]$. We may further suppose that $\mathfrak{D}(n)$ is invariant under ϕ as it is under P, with no other vectors in $\mathfrak{H}$ mapped into $\mathfrak{D}(n)$ by ϕ. Then the Lie expansion[15]

$$e^{i\phi}ne^{-i\phi} = n + i[\phi, n] + (i^2/2!)[\phi, [\phi, n]] + \cdots$$

is valid on the domain $\mathfrak{D}(n)$. We could then infer that the commutation relation $[n, \phi] = i1$ is equivalent to Eq. (3.9). Thus we are lead to attribute to ϕ the significance of a "phase angle," where $\hbar\phi$ is the observable canonically conjugate to n. The commutation relation in turn implies that the "phase angle" and number of quanta obey the Heisenberg uncertainty relation $\Delta\phi\Delta n \geq 1/2$. We must emphasize that no such self-adjoint operator does in fact exist, since its existence would imply the unitarity of P. However, the concepts introduced here become meaningful and useful under those limiting conditions where the operator P is approximately unitary. Moreover, the exponential serves as a heuristic device for suggesting other approaches to the problem of phase relations in the quantized radiation field.

A bounded operator P on $\mathfrak{H}$ can be written uniquely as the sum[16]

$$P = C + iS, \tag{3.14}$$

where C and S are bounded self-adjoint operators on $\mathfrak{H}$ given by

$$C = \tfrac{1}{2}(P + P^{\dagger}), \quad S = (1/2i)(P - P^{\dagger}). \tag{3.15}$$

The "real" and "imaginary" parts C and S commute if and only if P is a normal operator.

The symbols C and S are used for the components of our partial isometry P in Eq. (3.1) in analogy to the sine and cosine operators obtained as the corresponding components of a unitary operator. The commutation relations (3.11) and (3.12) give[6]

$$[C, n] = iS1, \quad [S, n] = -iC1, \tag{3.16}$$

and relations (3.3) give

$$[C, S] = \tfrac{1}{2}i\mathcal{P}, \tag{3.17a}$$

$$C^2 + S^2 = I - \tfrac{1}{2}\mathcal{P}. \tag{3.17b}$$

As expected, the presence of the ground state, which excludes the possibility of P being unitary (and hence normal), likewise prevents C and S from commuting. Measurements of the physical quantities C and S are compatible (approximately compatible) only in states of the radiation field which have no (negligible) ground-state component, in which case P is effectively unitary. The commutation relations (3.16) and (3.17a) imply

$$\Delta C\Delta n \geq \tfrac{1}{2}\langle S\rangle, \quad \Delta S\Delta n \geq \tfrac{1}{2}\langle C\rangle, \quad \Delta S\Delta C \geq \tfrac{1}{2}\langle\mathcal{O}\rangle, \quad (3.18)$$

which exhibit the mathematically rigorous forms of the number-phase uncertainty relations.

4. EIGENSTATES AND SPECTRA OF PHASE OPERATORS

A. Eigenstates and spectrum of P

In order to find possible eigenvectors of P and $P^\dagger$, i.e., possible solutions in $\mathfrak{H}$ of the eigenvalue equations $P\psi = u\psi$ and $P^\dagger\varphi = v\varphi$, where u and v are complex numbers, we write

$$P\psi = \sum_{n'=0}^{\infty} c_{n'+1}|n'\rangle = u\sum_{n'=0}^{\infty} c_{n'}|n'\rangle,$$

$$P^\dagger\varphi = \sum_{n'=1}^{\infty} b_{n'-1}|n'\rangle = v\sum_{n'=0}^{\infty} b_{n'}|n'\rangle.$$

Clearly, eigenvectors of P and $P^\dagger$ are determined by the solutions of the respective sets of equations

$$c_{n'+1} = uc_{n'}, \quad n' = 0, 1, 2, \cdots, \quad (4.1a)$$

$$b_0 = 0, b_{n'-1} = vb_{n'}, \quad n' = 1, 2, \cdots. \quad (4.1b)$$

The solutions of (4.1a) are given by $c_{n'} = cu^{n'}$. The coefficients satisfy $\sum|c_{n'}|^2 < \infty$ and the eigenvector belongs to $\mathfrak{H}$ if and only if $|u| < 1$. Therefore, in the complex u plane, the interior of the unit circle is the point spectrum of P and each eigenvalue is nondegenerate Thus, for a given mode of the radiation field, the states

$$|u\rangle = (1 - |u|^2)^{1/2}\sum_{n'=0}^{\infty} u^{n'}|n'\rangle, \quad |u| < 1, \quad (4.2)$$

are normalized eigenstates of the phase operator: $P|u\rangle = u|u\rangle$.

The norm of P is $|P| = 1$, and hence the spectrum of P must be contained in the closed unit circle $|u| \leq 1$. Moreover, P is a closed operator, so that its spectrum is a closed set. Since the interior of the unit circle is the point spectrum, the boundary $|u| = 1$ is the remainder of the spectrum. Because the interior contains all possible eigenvalues of P, for any value of u in $|u| \geq 1$, $P - uI$ is a one-to-one mapping of $\mathfrak{D}(P - uI) = \mathfrak{H}$ onto $\mathfrak{R}(P - uI)$. We now use the following theorem: If the inverse of a linear operator T exists, then T necessarily preserves the linear independence of an arbitrary set of vectors in its domain, i.e., the set $\{T\psi_n\}$ is linearly independent if and only if the set $\{\psi_n\}$ is linearly independent. From this it follows that, for $|u| = 1$, $P - uI$ transforms the basis $\{|n\rangle\}$ for $\mathfrak{H}$ into a linearly independent set contained in $\mathfrak{R}(P - uI)$. But the transformed set $\{|n-1\rangle - u|n\rangle\}$ is complete and hence the range $\mathfrak{R}(P - uI)$ is dense in $\mathfrak{H}$. Thus we conclude that P has the continuous spectrum $|u| = 1$ and has no residual spectrum.

The only solution of (4.1b) is the trivial one $b_n = 0$, $n = 0, 1, 2, \cdots$. Hence $P^\dagger$ has no point spectrum. Accordingly, for all values of v in the complex plane, $P^\dagger - vI$ is a one-to-one mapping of $\mathfrak{D}(P^\dagger - vI) = \mathfrak{H}$ onto $\mathfrak{R}(P^\dagger - vI)$. Now $|P^\dagger| = 1$, so that the spectrum of $P^\dagger$ is contained in $|v| \leq 1$. The basis $\{|n\rangle\}$ for $\mathfrak{H}$ is transformed by $P^\dagger - vI$ into the linearly independent set $\{e_n \equiv |n+1\rangle - v|n\rangle | n = 0, 1, 2, \cdots\}$, which is a (nonorthogonal) basis for for the subspace $\mathfrak{R}(P^\dagger - vI)$. Every vector φ in $\mathfrak{R}(P^\dagger - vI)$ has a unique expansion in this basis, given by the unique expansion in the basis $\{|n\rangle\}$ of the preimage of φ in $\mathfrak{H}$. The identity

$$|0\rangle - (1/v^N)|N\rangle = -\sum_{n=0}^{N-1} e_n/v^{n+1}$$

shows that the ground-state vector can be approached arbitrarily closely by some linear combination of the e_n only if $|v| > 1$.[17] Thus, for $|v| \leq 1$, the transformed basis is not complete in $\mathfrak{H}$ and $\mathfrak{R}(P^\dagger - vI)$ is not dense in $\mathfrak{H}$. It follows that $|v| \leq 1$ is the spectrum of $P^\dagger$ and is entirely a residual spectrum.

As we have already observed, the isometric transformation $P(P^\dagger)$ is maximal, inasmuch as its range (domain) is the entire Hilbert space $\mathfrak{H}$, but it is not unitary. Therefore neither P nor $P^\dagger$ possesses a resolution of the identity.[12] It follows that the phase eigenstates (4.2) of P do not form a complete set in $\mathfrak{H}$, for their completeness would imply that a resolution of the identity for P could be constructed as some linear combination of the projection operators $|u\rangle\langle u|$. An example of a complete set is given by the eigenstates of a, as shown by the spectral decomposition (2.18).

In a phase eigenstate $|u\rangle$, the expectation value of the photon number n for the given mode is

$$\langle n\rangle \equiv \langle u|n|u\rangle = (1 - |u|^2)\sum_{n'=0}^{\infty} n'|u|^{2n'} = |u|^2(1 - |u|^2)^{-1}, \quad (4.3)$$

and the expectation value of n^2 is

$$\langle n^2\rangle \equiv \langle u|n^2|u\rangle = (1 - |u|^2)\sum_{n'=0}^{\infty} n'^2|u|^{2n'}$$

$$= |u|^2(1 + |u|^2)(1 - |u|^2)^{-2}. \quad (4.4)$$

The dispersion in values of n for the phase eigenstate is the root-mean-square deviation:

$$\Delta n = \langle(n - \langle n\rangle)^2\rangle^{1/2} = (\langle n^2\rangle - \langle n\rangle^2)^{1/2} = |u|(1 - |u|^2)^{-1}. \quad (4.5)$$

Note that the state $|u = 0\rangle$ is equal to $|0\rangle$, the oscillator ground state. The phase of the vector $|n'\rangle$, $n' \neq 0$, is fixed relative to the ground state vector $|0\rangle$ by Eq. (2.9). Hence the choice of normalization constant in Eq. (4.2) to be the real number $(1 - |u|^2)^{1/2}$ determines the phase factor of the vector $|u\rangle$ relative to the ground state vector of the mode: $\langle u|0\rangle = (1 - |u|^2)^{1/2}$.

In the state $|u\rangle$, the probability of finding the oscillator in the energy state $|n'\rangle$ is

$$p(n') \equiv \langle u|n'\rangle\langle n'|u\rangle = (1 - |u|^2)|u|^{2n'}. \quad (4.6)$$

For a particular value of u, the distribution $p(n')$ has its maximum at $n' = 0$ and falls off exponentially with increasing n'. The distribution has the mean $\bar{n} = \sum_{n'=0}^{\infty} n'p(n') = \langle u|n|u\rangle = \langle n\rangle$ and variance $\sigma = [\sum_{n'=0}^{\infty}(n' - \bar{n})^2p(n')]^{1/2} = \Delta n$. Since $|u| < 1$ in Eqs. (4.3) and (4.5), the dispersion in values of n always exceeds the mean value of n. In this sense, the number of light quanta can be said to be undetermined in a phase eigenstate of a radiation oscillator. This statement is a

limited realization of the hypothetical uncertainty relation between the number operator n and a (nonexistent) phase angle observable ϕ.

When $|u| \lesssim 1$, the decrease of $p(n')$ with increasing n' is very slow. Thus for values of P near the unit circle, an eigenstate $|u\rangle$ has appreciable amplitudes for states $|n'\rangle$ with $n' \gg 1$ and the expectation value of n becomes very large. The contribution to $|u\rangle$ of the single state $|0\rangle$ then becomes relatively unimportant. Under these circumstances, we have seen in Sec. 3A that the operator P can be regarded as unitary and the classical description of the mode applies. Hence, the phase eigenstates have a classical electromagnetic field description when $|u| \lesssim 1$.

The limit $|u| = 1$ for a phase state (4.2), if it exists, implies that every value of n is equally probable. Let us define the vectors

$$|e^{i\theta}\rangle = (2\pi)^{-1/2} \sum_{n'=0}^{\infty} e^{in'\theta}|n'\rangle, \quad 0 \le \theta < 2\pi. \tag{4.7}$$

Such vectors do not exist in the Hilbert space $\mathfrak{H}$ since they do not have finite norm. They correspond to idealized states that are not physically realizable. Following the method of Dirac, $\mathfrak{H}$ can be extended to a vector space $\mathfrak{V}$ that contains the vectors $|e^{i\theta}\rangle$. From the definition (4.7), we see that the set $\{|n\rangle\}$ is also complete in $\mathfrak{V}$. The vectors (4.7) are not orthogonal. It can be shown[6] that

$$\langle e^{i\theta}|e^{i\theta'}\rangle = (4\pi)^{-1} + \tfrac{1}{2}\delta(\theta - \theta') - (i/4\pi)\cot\tfrac{1}{2}(\theta - \theta').$$

However, the vectors (4.7) form a complete set for $\mathfrak{V}$, inasmuch as

$$\int_0^{2\pi} d\theta\, |e^{i\theta}\rangle\langle e^{i\theta}| = \sum_{n',n''=0}^{\infty} |n'\rangle\langle n''| \int_0^{2\pi} e^{i\theta(n''-n')}\, d\theta/2\pi$$
$$= \sum_{n'=0}^{\infty} |n'\rangle\langle n'| = \mathcal{I},$$

where $\mathcal{I}$ is the identity operator in $\mathfrak{V}$.

B. Spectrum of C and of S

According to the definitions (3.15) of C and S, their norms satisfy

$$|C| \le \tfrac{1}{2}(|P| + |P^\dagger|) = 1, \quad |S| \le \tfrac{1}{2}(|P| + |P^\dagger|) = 1.$$

It turns out that $|C| = |S| = 1$. Since $C(S)$ is self-adjoint, its spectrum lies on the real axis in the interval $[-1, 1]$ and contains no residual part.

It is easily shown that $C(S)$ has no point spectrum. Let $|c'\rangle$ denote an eigenvector of C belonging to the eigenvalue c', and expand $|c'\rangle$ in the complete set of energy eigenstates:

$$|c'\rangle = \sum_{n'=0}^{\infty} d_{n'}|n'\rangle. \tag{4.8}$$

The eigenvalue equation $C|c'\rangle = c'|c'\rangle$ is equivalent to the set of recursion relations

$$d_1 = 2c'd_0, \quad d_{n+2} = 2c'd_{n+1} - d_n. \tag{4.9}$$

The solution of this set of equations is given by

$$d_n = dz^n + bz^{-n}, \quad n \ge 2, \tag{4.10}$$

provided $2c' = z + z^{-1}$. Then the eigenvalue c' will be a real number if and only if $|z| = 1$. But when z lies on the unit circle, d_n does not approach zero, in the limit $n \to \infty$, for any set of values of the constants d, b, and the series (4.8) cannot converge to a vector in $\mathfrak{H}$. Thus C has no eigenvectors in $\mathfrak{H}$, i.e., C has no point spectrum.

The spectrum of C must be purely continuous. The easiest way to treat the continuous spectrum of a self-adjoint operator is the method of Dirac. Since the reality of c' requires that $|z| = 1$, we write $z = e^{i\theta}$ and have $c' = \cos\theta$. We next observe that the relations (4.9) imply $d_n = d_0 r_n$ for $n \ge 1$, where r_n is a real number. This condition is satisfied if $d = d_0 h$, $b = d_0 h^*$ in Eq. (4.10). If we write $d = K(e^{i\theta}/2i)$, $b = K(-e^{-i\theta}/2i)$, then $d_n = K\sin(n+1)\theta$, $n \ge 2$. The choice $d_0 = K\sin\theta$ makes the expression for d_n hold for all values of n, as may be verified directly from the Eqs. (4.9). With the normalization $K = (2/\pi)^{1/2}$, the eigenvectors[5]

$$|\cos\theta\rangle = (2/\pi)^{1/2} \sum_{n'=0}^{\infty} \sin(n'+1)\theta\,|n\rangle, \quad 0 \le \theta \le \pi, \tag{4.11}$$

satisfy the orthonormality condition

$$\langle\cos\theta|\,\cos\theta'\rangle = \delta(\theta - \theta'), \quad \theta \ne 0, \pi, \quad \theta' \ne 0, \pi.$$

The interval $0 \le \theta \le \pi$ gives all the independent eigenvectors, since $|\cos(-\theta)\rangle = -|\cos\theta\rangle$. A single independent eigenvector belongs to each value of θ inside the interval. Thus the spectrum of C is continuous and consists of the entire closed interval $[-1, 1]$.

If we extend the Hilbert space $\mathfrak{H}$ to a vector space $\mathfrak{V}_c$ containing the eigenvectors of C, we see from Eq. (4.11) that the set $\{|n'\rangle\}$ is a basis for $\mathfrak{V}_c$. The set of eigenvectors form a complete orthonormal set in $\mathfrak{V}_c$, since

$$\int_0^{\pi} d\theta\,|\cos\theta\rangle\langle\cos\theta| = (2/\pi)\sum_{n',n''} |n'\rangle\langle n''| \int_0^{\pi} d\theta\, \sin(n'+1)\theta$$
$$\times \sin(n''+1)\theta = \sum_{n'} |n'\rangle\langle n'| = \mathcal{I}_c,$$

where $\mathcal{I}_c$ is the identity operator in $\mathfrak{V}_c$. The relation

$$|\cos\theta\rangle = -i(e^{i\theta}|e^{i\theta}\rangle - e^{-i\theta}|e^{-i\theta}\rangle)$$

shows $\mathfrak{V}_c$ is a subspace of the vector space $\mathfrak{V}$ defined in Sec. 4A.

Since S and C do not commute, they do not possess a common set of eigenvectors. The eigenvectors and spectrum of S are obtained by an essentially similar development. The spectrum of S is purely continuous in the entire interval $[-1, 1]$, and to each value in the interval there corresponds one independent eigenvector The eigenvectors of S form a complete orthonormal set in a vector space $\mathfrak{V}_s$, which is an extension of $\mathfrak{H}$ and a subspace of $\mathfrak{V}$. Generalized "cosine" and "sine" operators, as well as annihilation type operators corresponding to pairs of such operators, have been studied by Lerner, *et al.*[18] The properties of these operators are further elucidated by Aharonov, *et al.*[19]

5. RELATIVE PHASE OPERATORS

We now investigate the generalization to quantized fields of the relative phase (phase difference) between a pair of modes. The discussion involves two distinct modes, and we shall accordingly revert to the modal index notation. Although the symbolic expression $e^{i\phi k}$ for the phase operator P_k has meaning only in the classical limit, the analogous exponential function $\exp[i(\phi_k - \phi_{k'})]$ of the relative phase angle suggests that we consider the operator $P_{k'}^\dagger P_k$, $k' \ne k$.

Such mixed operator products as $P_{k'}^\dagger P_k$ and $P_{k'} P_k^\dagger$,

$k' \neq k$, are defined on the Hilbert space $\mathfrak{H}_{k'k} \equiv \mathfrak{H}_{k'} \otimes \mathfrak{H}_k$. Because of the kinematical independence of the different modes, any two operators pertaining to distinct modes commute, e.g.,

$$[P^\dagger_{k'}, P_k] = [P_{k'}, n_k^{1/2}] = 0, \quad k' \neq k. \tag{5.1a}$$

Likewise, it is clear that

$$[P_{k'}, P_k] = [n_{k'}^{1/2}, n_k^{1/2}] = 0, \quad \text{any pair } k, k'. \tag{5.1b}$$

The operator $P^\dagger_{k'} P_k$ is bounded on $\mathfrak{H}_{k'} \otimes \mathfrak{H}_k$ and can be written uniquely as the sum

$$P^\dagger_{k'} P_k = C_{k'k} + iS_{k'k},$$

where $C_{k'k}$ and $S_{k'k}$ are bounded self-adjoint operators on $\mathfrak{H}_{k'k}$ given by

$$C_{k'k} = \tfrac{1}{2}(P^\dagger_{k'} P_k + P^\dagger_k P_{k'}) = C_{kk'}, \quad k' \neq k, \tag{5.2a}$$

$$S_{k'k} = (1/2i)(P^\dagger_{k'} P_k - P^\dagger_k P_{k'}) = -S_{kk'}, \quad k' \neq k. \tag{5.2b}$$

Using relations (3.3) and (3.4), we find

$$[C_{k'k}, S_{k'k}] = (1/2i)(\mathcal{P}_k - \mathcal{P}_{k'}). \tag{5.3}$$

The total photon number for the modes k', k is

$$N = n_{k'} I_k + I_{k'} n_k \equiv n_{k'} + n_k. \tag{5.4}$$

The domain of N is the subspace $\mathfrak{D}(n_{k'}) \otimes \mathfrak{D}(n_k)$ of $\mathfrak{H}_{k'k}$. With the help of the commutation relations (3.16), a short calculation shows that N commutes with both $C_{k'k}$ and $S_{k'k}$:

$$[C_{k'k}, N] = 0, [S_{k'k}, N] = 0, \tag{5.6}$$

where it is understood the domain of each commutator is $\mathfrak{D}(N)$. This means that a relative phase operator can be measured simultaneously with the total number of photons in the two modes (but not with the number of photons in each individual mode). Equivalently, a common set of eigenvectors can be chosen for $C_{k'k}$ and N and for $S_{k'k}$ and N. Because of (5.3), however, there exists no common set of eigenvectors for all three operators. Since $\mathcal{P}_k n_k$ is equal to the zero operator on $\mathfrak{D}(n_k)$ and $n_k \mathcal{P}_k = 0$, N also commutes with the commutator (5.3).

The state $|n'_{k'}\rangle |n'_k\rangle \equiv |n'_{k'}, n'_k\rangle$ is an eigenstate of N with the eigenvalue $N' = n'_{k'} + n'_k$, and the set of all such vectors is a complete orthonormal basis for $\mathfrak{H}_{k'k}$. Let $\mathfrak{H}_{N'}$ be the subspace spanned by all the basic product states having a total of N' photons in modes k' and k. Clearly $\mathfrak{H}_{N'}$ has dimension $N' + 1$ and consists of all eigenvectors of N with the eigenvalue N' and of only such vectors. Any vector in $\mathfrak{H}_{N'}$ has the form

$$\psi_{N'} = \sum_{n'=0}^{N'} g_{n'} |N' - n', n'\rangle.$$

The Hilbert space $\mathfrak{H}_{k'k}$ is the direct sum of the orthogonal subspaces $\mathfrak{H}_{N'}$, $N' = 0, 1, 2, \cdots$:

$$\mathfrak{H}_{k'k} = \bigoplus_{N'} \mathfrak{H}_{N'}.$$

The commutativity of the relative phase operators with N means that each subspace $\mathfrak{H}_{N'}$ is invariant under $C_{k'k}$ and $S_{k'k}$. Let $|N'; c'\rangle$ denote a simultaneous eigenvector of N and $C_{k'k}$ with the indicated eigenvalues. For $N' = 0$, the basic state $|0, 0\rangle$ is the eigenstate of $C_{k'k}$ and has the eigenvalue zero. In the one photon subspace, which is two-dimensional, $C_{k'k}$ has two orthogonal eigenvectors, which can be taken as

$$|1; \pm 1/2\rangle = (2)^{-1/2}[|1, 0\rangle \pm |0, 1\rangle].$$

In the three-dimensional subspace $\mathfrak{H}_2$ of two-photon states, $C_{k'k}$ has three orthogonal eigenvectors with the eigenvalues $0, \pm (2)^{-1/2}$. The generalization to arbitrary $\mathfrak{H}_{N'}$ is as follows[6]: $C_{k'k}$ has the $N' + 1$ eigenvalues

$$c'_r = \cos\theta_{N'r}, \quad \theta_{N'r} = \pi r/(N' + 2),$$
$$r = 1, 2, \ldots, N' + 1, \tag{5.7}$$

and the corresponding eigenvectors

$$|N'; \cos\theta_{N'r}\rangle = (2/N' + 2)^{1/2} \sum_{n'=0}^{N'} \sin(n' + 1)\theta_{N'r} |n', N' - n'\rangle \tag{5.8}$$

form a complete orthonormal set in $\mathfrak{H}_{N'}$. The $N' + 1$ angles $\theta_{N'r}$ are uniformly distributed in the interval $(0, \pi)$. It is evident that the set of angles given by all values of N' is denumerable and dense in the interval. The set of all angles is in one-to-one correspondence with the eigenvalues, which are seen to form a denumerable and dense set of points on the real line in the interval $(-1, 1)$. Moreover, the collection of all eigenvectors (5.8) is a complete orthonormal set in $\mathfrak{H}_{k'k}$. Hence we conclude that $C_{k'k}$ has a pure point spectrum that is everywhere dense in the interval $[-1, 1]$. The spectral decomposition of $C_{k'k}$ is

$$C_{k'k} = \sum_{N'=0}^{\infty} \sum_{r=1}^{N'+1} \cos\theta_{N'r} |N'; \cos\theta_{N'r}\rangle\langle N'; \cos\theta_{N'r}|.$$

The relative phase operator for any pair of modes k', k may be regarded as an operator in the Fock space $\oplus_n \mathfrak{H}^{(n)}$, where $\mathfrak{H}^{(n)}$ is the subspace of all n-photon states. The states (5.8) are the N'-photon states for the two modes k' and k. Let $\overline{\mathfrak{H}}_n$ be the subspace of n-photon states for the collection of all modes except k' and k. Then $\mathfrak{H}^{(n)}$ can be decomposed into the direct sum of orthogonal subspaces

$$\mathfrak{H}^{(n)} = \bigoplus_{N'=0}^{n} \mathfrak{H}_{N'} \otimes \overline{\mathfrak{H}}_{n-N'},$$

where the subspace $\mathfrak{H}_{N'} \otimes \overline{\mathfrak{H}}_{n-N'}$ consists of all n-photon states which are eigenstates of $C_{k'k}$ with eigenvalues (5.7) given by the indicated value of N. Clearly $C_{k'k}$ has a complete set of eigenstates in $\mathfrak{H}^{(n)}$, $n = 0, 1, 2, \cdots$. Thus as an operator in the separable Fock space, the spectrum of a relative phase operator $C_{k'k}$ is the same as its restriction to the subspace $\mathfrak{H}_{k'k}$.

The spectrum and eigenvectors of $S_{k'k}$ are obtained by a development essentially parallel to that for $C_{k'k}$. The eigenvalues of $S_{k'k}$ are given by $s'_r = \sin\varphi_{N'r}$, $\varphi_{N'r} = [\pi r/(N' + 2)] - \pi/2$, $r = 1, 2, \ldots, N' + 1$.

6. PHASE RELATIONS IN PHOTON STATES

A. The phase operators for a single mode

First we investigate the values of the self-adjoint phase operators C and S when the mode is in a normalizable phase state (4.2). Only a single mode enters the

discussion, and so the modal subscript can be omitted. The phase state $|u\rangle$ is an eigenvector of the phase operator P, and hence

$$P^n|u\rangle = u^n, \quad \langle u|(P^\dagger)^n = (u^*)^n. \tag{6.1}$$

The expectation values of C and S follow immediately:

$$\langle u|C|u\rangle = \tfrac{1}{2}(u + u^*) = \mathcal{R}e(u),$$
$$\langle u|S|u\rangle = (1/2i)(u - u^*) = \mathcal{I}m(u). \tag{6.2}$$

With the help of Eqs. (6.1) and (3.3), we find

$$\langle u|C^2|u\rangle = \tfrac{1}{4}(u^2 + u^{*2} + |u|^2 + 1),$$
$$\langle u|S^2|u\rangle = -\tfrac{1}{4}(u^2 + u^{*2} - |u|^2 - 1),$$

which combine with relations (6.2) to give the dispersion in values of C and of S:

$$\Delta C = (\langle u|C^2|u\rangle - \langle u|C|u\rangle^2)^{1/2} = \tfrac{1}{2}(1 - |u|^2)^{1/2}, \tag{6.3a}$$

$$\Delta S = (\langle u|S^2|u\rangle - \langle u|S|u\rangle^2)^{1/2} = \tfrac{1}{2}(1 - |u|^2)^{1/2}. \tag{6.3b}$$

As expected in the classical domain $|u| \approx 1$, the dispersions become negligible and the phase observables have well determined values. The states $|u\rangle$ behave as eigenstates of C and S in the classical limit. Whenever the classical description applies, the mean value of n is large and the value of n is highly indeterminate. According to the uncertainty relations (3.18), a large value of Δn is necessary in order that the value of a phase observable be well determined. Equations (4.5) and (6.3a) give the explicit result

$$\Delta C\, \Delta n = \tfrac{1}{2}|u|(1 - |u|^2)^{-1/2}. \tag{6.4}$$

Thus, in the phase state $|u\rangle$, the product of the uncertainties in C and in n becomes very large in the classical domain.

The values of C and S and the number-phase uncertainty relations for the coherent states have been discussed in detail by Carruthers and Nieto.[6] We sketch the results for comparison with the phase states. From the relations

$$P|\alpha\rangle = \alpha e^{-(1/2)|\alpha|^2} \sum_{n=0}^{\infty} [(n+1)!]^{-1/2}\alpha^n |n\rangle,$$
$$P^2|\alpha\rangle = \alpha^2 e^{-(1/2)|\alpha|^2} \sum_{n=0}^{\infty} [(n+2)!]^{-1/2}\alpha^n |n\rangle,$$

we obtain

$$\langle\alpha|P|\alpha\rangle = \alpha e^{-|\alpha|^2}F_1(|\alpha|^2),$$
$$\langle\alpha|P^2|\alpha\rangle = \alpha^2 e^{-|\alpha|^2}F_2(|\alpha|^2), \tag{6.5}$$

where the functions F_1 and F_2 are defined by the power series

$$F_1(z) = \sum_{n=0}^{\infty} \frac{z^n}{n!(n+1)^{1/2}},$$
$$F_2(z) = \sum_{n=0}^{\infty} \frac{z^n}{n![(n+1)(n+2)]^{1/2}}. \tag{6.6}$$

Since both series converge for all values of z in the complex plane, they define entire functions F_1 and F_2. In terms of these functions, we find

$$\langle\alpha|C|\alpha\rangle = \tfrac{1}{2}(\alpha + \alpha^*)e^{-|\alpha|^2}F_1(|\alpha|^2),$$
$$\langle\alpha|S|\alpha\rangle = (1/2i)(\alpha - \alpha^*)e^{-|\alpha|^2}F_1(|\alpha|^2), \tag{6.7}$$

$$\langle\alpha|C^2|\alpha\rangle = \tfrac{1}{4}[(\alpha^2 + \alpha^{*2})e^{-|\alpha|^2}F_2(|\alpha|^2) + 2 - e^{-|\alpha|^2}]. \tag{6.8}$$

The asymptotic representations

$$F_1(z) \sim z^{-1/2}e^z[1 - (1/8z) + \cdots],$$
$$F_2(z) \sim z^{-1}e^z[1 - (1/2z) + \cdots] \quad (|z| \gg 1), \tag{6.9}$$

give

$$\langle\alpha|P|\alpha\rangle \sim \alpha/|\alpha|, \quad \Delta C \sim 0 \quad (|\alpha| \gg 1). \tag{6.10}$$

Thus, for coherent states in the classical domain, the expectation value of P shows the anticipated character of a phase factor and the dispersion in values of C becomes negligible. However, with the help of Eq. (2.14), the product of the uncertainties is found to have the asymptotic behavior $\Delta C\, \Delta n \sim \frac{1}{2}\langle S\rangle$. Hence in the classical domain, the coherent states behave as minimum-uncertainty states for the number-phase uncertainty relations, in contrast to the behavior (6.4) of the phase states. Similar results are obtained with the phase operator S.

The situation in which $\langle n\rangle \ll 1$ is sometimes called the "quantum limit." Observe that the dispersion in the phase operator attains its largest value in this limiting case.

B. The relative phase operators

With the help of relations (5.1a) and (3.3), we write

$$C_{k'k}^2 = \tfrac{1}{4}(P_{k'}^{\dagger 2}P_k^2 + P_k^{\dagger 2}P_{k'}^2 + P_{k'}^\dagger P_{k'} I_k + P_k^\dagger P_k I_{k'}).$$

Let us first consider the case where the state of the radiation field is such that the modes k' and k occupy phase eigenstates of their respective phase operators, $P_{k'}$ and P_k. We denote the state vector in $\mathfrak{H}_{k'} \otimes \mathfrak{H}_k$ describing the two modes by $|u_{k'}, u_k\rangle \equiv |u_{k'}\rangle_{k'}|u_k\rangle_k$. From Eqs. (6.1) the expectation value of the relative phase operator and its square follow directly:

$$\langle C_{k'k}\rangle \equiv \langle u_{k'}, u_k|C_{k'k}|u_{k'}, u_k\rangle = \tfrac{1}{2}(u_{k'}^* u_k + u_{k'} u_k^*),$$
$$\langle C_{k'k}^2\rangle = \tfrac{1}{4}[(u_{k'}^* u_k)^2 + (u_{k'} u_k^*)^2 + |u_{k'}|^2 + |u_k|^2].$$

The uncertainty in values of the relative phase when the two modes are in phase eigenstates is

$$\Delta C_{k'k} = (\langle C_{k'k}^2\rangle - \langle C_{k'k}\rangle^2)^{1/2} = \tfrac{1}{2}[|u_{k'}|^2(1 - |u_k|^2) + |u_k|^2(1 - |u_{k'}|^2)]^{1/2}. \tag{6.11}$$

The relative phase has a well-defined value only if (i) $|u_k| \approx 0$, $|u_{k'}| \approx 0$ or if (ii) $|u_k| \approx 1$, $|u_{k'}| \approx 1$. Condition (i) is the quantum limit and corresponds to the situation where both oscillators are essentially in their ground states. Recall that the state $u_k = u_{k'} = 0$, with both oscillators in their ground state, is an eigenstate of the relative phase. Condition (ii) represents the other extreme from (i), the limit in which both modes have a classical description.

Next consider the case where the modes k and k' are in coherent states. Let $|\alpha_{k'}, \alpha_k\rangle \equiv |\alpha_{k'}\rangle_{k'}|\alpha_k\rangle_k$ be the state vector describing the two modes. A short calculation with the help of Eqs. (6.5) gives

$$\langle\alpha_{k'}, \alpha_k|C_{k'k}|\alpha_{k'}, \alpha_k\rangle = \tfrac{1}{2}(\alpha_k^*\alpha_k + \alpha_{k'}\alpha_k^* \times \exp(-|\alpha_{k'}|^2 - |\alpha_k|^2)F_1(|\alpha_{k'}|^2)F_1(|\alpha_k|^2), \tag{6.12}$$

$$\langle \alpha_{k'}, \alpha_k | C^2_{k'k} | \alpha_{k'}, \alpha_k \rangle$$
$$= \tfrac{1}{2}[1 - \tfrac{1}{2}\exp(-|\alpha_{k'}|^2) - \tfrac{1}{2}\exp(-|\alpha_k|^2)]$$
$$+ \tfrac{1}{4}(\alpha^2_{k'}\alpha^{*2}_k + \alpha^{*2}_{k'}\alpha^2_k)\exp(-|\alpha_{k'}|^2 - |\alpha_k|^2)$$
$$\times F_2(|\alpha_{k'}|^2)F_2(|\alpha_k|^2). \quad (6.13)$$

Again the uncertainty in the relative phase is small in the two extremes: (1) the quantum limit $\langle n \rangle = |\alpha_k|^2 \ll 1$, $\langle n_{k'} \rangle = |\alpha_{k'}|^2 \ll 1$ and (2) the classical limit $|\alpha_k| \gg 1$, $|\alpha_{k'}| \gg 1$. In the quantum limit, the state of the two modes is essentially the relative phase eigenstate $|0, 0\rangle$, as already noted.

The form of Eqs. (6.12) and (6.13) in the classical domain derives from the asymptotic expansions (6.9). If the numbers α are expressed in polar form, $\alpha = |\alpha| e^{i\varphi}$, the classical case is summarized by

$$\langle \alpha_{k'}, \alpha_k | C_{k'k} | \alpha_{k'}, \alpha_k \rangle \approx \cos(\varphi_k - \varphi_{k'})\Delta C_{k'k} \approx 0,$$
$$(|\alpha_k| \gg 1, |\alpha_{k'}| \gg 1), \quad (6.14)$$

as expected.

We conclude that the classical field concept of a coherent electromagnetic wave having a well-defined phase for each modal component and a well-defined relative phase between each pair of modes is realized only in the classical domain. The role of phase operators in describing photon states is subject to this limitation.

C. Values of the phase operators

In Eq. (2.2) a change in the origin of the time parameter is equivalent to a change in each amplitude operator by a constant phase factor. Similarly, when the mode functions are traveling waves, a change in the origin of the spatial reference axes corresponds simply to a new phase factor in each modal amplitude. These observations are consistent with the fact that the commutation relations (2.6) are invariant under any transformation of the form

$$a_k \to \bar{a}_k = e^{i\gamma_k} a_k, \qquad a^\dagger_k \to \bar{a}^\dagger_k = e^{-i\gamma_k} a^\dagger_k, \quad (6.15)$$

where γ_k is a real number.

We see that for a given radiation field, the amplitude operators remain undetermined to within constant phase factors until the space–time origin, phases of the modal wavefunctions, etc., are prescribed. A corresponding constant phase factor remains to be fixed in the values of each modal phase operator. According to Eq. (2.9), the transformation (6.15) changes the phase of an energy eigenstate vector:

$$|n'_k\rangle \to \overline{|n'_k\rangle} = e^{-in'_k\gamma_k}|n'_k\rangle. \quad (6.16)$$

Recall that the phase factor of every energy eigenvector $|n_k\rangle$ is referred to the ground state vector $|0\rangle_k$. A change in phase of $|0\rangle_k$ merely introduces an equal phase change in all the energy eigenvectors.

In a phase eigenstate (4.2)

$$|u_k\rangle = (1 - |u_k|^2)^{1/2} \sum_{n'} u^{n'}_k |n'\rangle_k$$
$$= (1 - |\bar{u}_k|^2)^{1/2} \sum_{n'} \bar{u}^{n'}_k \overline{|n'\rangle}_k,$$

the transformation (6.16) is seen to be equivalent to a change in the phase value

$$u_k \to \bar{u}_k = e^{i\gamma_k} u_k. \quad (6.17)$$

The transformation (6.15) induces a corresponding transformation in the phase operator:

$$P_k \to \overline{P}_k = e^{i\gamma_k} P_k. \quad (6.18)$$

The phase state $|u_k\rangle$ is an eigenstate of $\overline{P}_k$ with eigenvalue $\bar{u}_k$, as a comparison of Eqs. (6.17) and (6.18) immediately reveals. Similarly, under (6.15) the relative phase operator transforms as

$$C_{k'k} \to \overline{C}_{k'k} = \tfrac{1}{2}(e^{i(\gamma_k - \gamma_{k'})} P^\dagger_k P_k + e^{-i(\gamma_k - \gamma_{k'})} P^\dagger_{k'} P_{k'}),$$

with an obvious change in the values of the relative phase.

[1] R. J. Glauber, Phys. Rev. 130, 2529 (1963).
[2] R. J. Glauber, Phys. Rev. 131, 2766 (1963).
[3] P. A. M. Dirac, Proc. Roy. Soc. (London) A 114, 243 (1927).
[4] W. Heitler, *The Quantum Theory of Radiation* (Oxford U.P., London, 1954), 3rd ed., p. 65.
[5] L. Susskind and J. Glogower, Physics 1, 49 (1964).
[6] P. Carruthers and M. M. Nieto, Rev. Mod. Phys. 40, 411 (1968).
[7] R. F. Streater and A. S. Wightman, *PCT, Spin and Statistics, and All That* (Benjamin, New York, 1964), p. 87.
[8] N. Dunford and J. T. Schwartz, *Linear Operators* (Wiley, New York, 1963), Part II, pp. 1248–50.
[9] M. Reed and B. Simon, *Methods of Modern Mathematical Physics* (Academic, New York, 1972), Part 1, pp. 297-98.
[10] N. Dunford and J. T. Schwartz, *Linear Operators* (Wiley, New York, 1971), Part III, Chap. XVIII.
[11] F. Riesz and B. Sz.-Nagy, *Functional Analysis* (Ungar, New York, 1955), 2nd ed., p. 281.
[12] J. von Neumann, *Mathematical Foundations of Quantum Mechanics* (Princeton U.P., Princeton, N.J., 1955), pp. 167–8.
[13] See Ref. 8, especially Chap. XII, and Ref. 10, especially Chap. XVIII.
[14] See Ref. 11, especially Chap. VIII.
[15] A. Messiah, *Quantum Mechanics* (North-Holland, Amsterdam, 1961), Vol. 1, p. 339.
[16] See Ref. 8, p. 906.
[17] The same conclusion holds for any energy state $|n'\rangle$, as can be seen by taking appropriate linear combinations of the e_n.
[18] E. C. Lerner, H. W. Huang, and G. E. Walters, J. Math. Phys. 11, 1679. (1970).
[19] Y. Aharonov, E. C. Lerner, H. W. Huang, and J. M. Knight, J. Math. Phys. 14, 746 (1973).

Invariant trilinear couplings involving both $SU(2)$ and $SU(1,1)$ states

J. Patera and P. Winternitz

Centre de Recherches Mathématiques, Université de Montréal, Montréal 101, P.Q., Canada
(Received 19 June 1973)

Multiplets $|JM>$ transforming under irreducible representations of the algebra of $SU(2)$ are constructed as infinite sums of products of two $SU(1,1)$ states $|'_{-p}>$ and $|'_{M+p}>$. The "generalized coupling coefficients" figuring in the construction are shown to exist if the $SU(2)$ label J and the $SU(1,1)$ labels j and l satisfy $j = l + J - k$, $0 \leq k \leq 2J$. They are constructed explicitly and turn out to be analytic continuations of both $SU(2)$ and $SU(1,1)$ Clebsch–Gordan coefficients. The constructed states $|JM>$ are not normalizable in the usual sense. The "generalized coupling coefficients" can be applied, e.g., to study vertices, involving ordinary particles (with mass satisfying $m^2 > 0$ and $SU(2)$ spin) and tachyons ($m^2 < 0$, $SU(1,1)$ spin).

1. INTRODUCTION

The purpose of this article is to consider a somewhat unusual mathematical object, namely a generalization of Clebsch–Gordan coefficients, making it possible to construct $SU(2)$ multiplets (in some generalized sense) out of products of $SU(1,1)$ multiplets. Indeed, let us write down the identity

$$|JMjl\rangle \equiv |JM\rangle \equiv \sum_{m+n=M} \begin{pmatrix} j & l & J \\ n & m & M \end{pmatrix} \left| \begin{matrix} j \\ n \end{matrix} \right\rangle \left| \begin{matrix} l \\ m \end{matrix} \right\rangle, \qquad (1)$$

where $|{}^{j}_{n}\rangle$ and $|{}^{l}_{m}\rangle$ transform according to irreducible unitary representations of the group $SU(1,1)$. The question we raise is: Is it possible to construct such "generalized Clebsch–Gordan coefficients" $\left(\begin{smallmatrix} j & l & J \\ n & m & M \end{smallmatrix}\right)$ that the object $|JM\rangle$ on the left-hand side of (1) transforms according to an irreducible unitary (and finite-dimensional) representation of $SU(2)$ (so that J is integer or half-integer and $M = -J, -J+1, \ldots, J$). The group $SU(1,1) \times SU(1,1)$ does not contain an $SU(2)$ subgroup, hence a direct product of $SU(1,1)$ representations cannot be decomposed into $SU(2)$ ones. The vectors $|JM\rangle$ in (1), if they exist, can thus not be vectors in the direct product Hilbert space of two $SU(1,1)$ representations.

In this article we shall however demonstrate that if the numbers j and l labeling $SU(1,1)$ representations and J, labeling $SU(2)$ ones, satisfy the condition

$$j = l + J - k, \qquad 0 \leq k \leq 2J, \qquad (2)$$

then such coefficients $\left(\begin{smallmatrix} j & l & J \\ n & m & M \end{smallmatrix}\right)$ can be found that the vector $|JM\rangle$ transforms like a basis vector for an irreducible unitary representation of the algebra of $SU(2)$, i.e., satisfies

$$\begin{aligned} J_{\pm}|JM\rangle &= [(J \mp M)(J \pm M + 1)]^{1/2}|JM \pm 1\rangle, \\ J_3|JM\rangle &= M|JM\rangle, \end{aligned} \qquad (3)$$

where $J = 0, \frac{1}{2}, 1, \cdots$ and $M = -J, -J+1, \cdots, J$ [$J_{\pm}$ and J_3 are generators of $SU(2)$, specified below]. The basis vectors defined by (1) are, however, not normalizable and satisfy

$$\langle J'M'l'j'|JMlj\rangle = \begin{cases} \infty & \text{if } M = M',\ j = j' \text{ and } l = l' \\ 0 & \text{otherwise} \end{cases} \qquad (4)$$

regardless of the values of J and J'. Note that the action of the operators $J_{\pm}$ and J_3 on $|JM\rangle$ is defined by Eq. (1) [their action on the right-hand side of (1) is known], and that the multiplets $|JM\rangle$ are defined up to an arbitrary complex factor, depending on J, j and l, but not on M.

In this article we shall not discuss the details of any physical applications of the presented "generalized Clebsch–Gordan coefficients" or "generalized coupling coefficients." By way of motivation let us, however, mention that these objects will occur if we consider vertices, involving one tardyon [a usual particle with positive mass squared $m^2 > 0$ and spin described[1] by the group $SU(2)$], and two tachyons[2] [faster than light particles with $m^2 < 0$ and spin described by the group $SU(1,1)$]. Similarly, Regge pole theory can be interpreted as an expansion in terms of irreducible representations of the group $SU(1,1)$ (figuring as the little group[1] of the Poincaré group, leaving a spacelike vector, namely the momentum transfer, invariant). Regge poles then correspond to certain nonunitary infinite-dimensional irreducible representations of $SU(1,1)$, the background integral corresponds to unitary representations of the principal series, and "nonsense terms" correspond to unitary representations of the discrete series.[3] The coupling constant in (1) can then be interpreted in terms of vertices involving particles and reggeons (or other objects figuring in complex angular momentum theory). Finally, there are quite a few other areas of physics in which the representations of $SU(1,1)$ play an important role, such as the theory of the hydrogen atom,[4] various internal symmetries in particle physics,[5] etc. The generalized coupling coefficients of this article relating $SU(2)$ and $SU(1,1)$ multiplets may well prove to be useful in any of the above areas, as well as other ones.

In Sec. 2 of this article we give, mainly for purposes of notation, a brief review of the representation theory of the $SU(1,1)$ and $SU(2)$ groups. In Sec. 3 we derive expressions for the coupling coefficients $\left(\begin{smallmatrix} j & l & J \\ \pm J+p & -p & \pm J \end{smallmatrix}\right)$ corresponding to the highest and lowest weights of $SU(2)$ representations. The general coupling coefficients as well as selection rules for the values of j, l and J are presented in Sec. 4. Special cases for low values of J are discussed in Sec. 5. In Sec. 6 we discuss various properties of the generalized coupling coefficients, in particular their normalization, symmetries, recursion relations, etc. We demonstrate that the coupling coefficients of this article can be considered to be analytic continuations of both the $SU(2)$[6] and the $SU(1,1)$[7] Clebsch–Gordan coefficients. We show that it is possible to construct $SU(1,1)$ multiplets out of direct products of $SU(2)$ and $SU(1,1)$ ones, but that the couplings $SU(2) \times SU(2) \to SU(1,1)$ and $SU(2) \times SU(1,1) \to SU(2)$ are not possible. In the final Sec. 7 we make some further comments on the significance and applicability of the results of this paper and discuss possible future developments.

2. REPRESENTATIONS OF THE $SU(2)$ AND $SU(1,1)$ ALGEBRAS

Let us spell out some well-known results in order to establish notation. The algebra of the group $SU(2)$ [or the

locally isomorphic rotation group $O(3)$] can be so defined that the generators L_i satisfy the commutation relations

$$[L_i, L_k] = \epsilon_{ikl} L_l \quad (i,k,l = 1,2,3). \tag{5}$$

Similarly, the algebras of $SU(1,1)$ [or the locally isomorphic three-dimensional Lorentz group $O(2,1)$] can be defined so that the algebra satisfies

$$[K_1, K_2] = -L_3, \ [L_3, K_1] = K_2, \ [K_2, L_3] = K_1. \tag{6}$$

With the conventions (5) and (6) all operators in a unitary representation of the corresponding algebra will be anti-Hermitian.

For our purposes it is more convenient to define different bases for the algebras of $SU(2)$ and $SU(1,1)$, namely such that the matrix elements of the generators in irreducible representations have the same form for both algebras. Thus, we introduce the standard raising and lowering operators (see, e.g., Ref. 8) and for $SU(2)$ put,

$$J_+ = -iL_1 + L_2, \ J_- = -iL_1 - L_2, \quad J_3 = iL_3, \tag{7}$$

whereas for $SU(1,1)$ we put

$$J_+ = K_1 + iK_2, \quad J_- = K_1 - iK_2, \quad J_3 = iL_3. \tag{8}$$

The commutation relations for $SU(2)$ and $SU(1,1)$ can both be written as

$$[J_3, J_+] = J_+, \quad [J_3, J_-] = -J_-, \quad [J_+, J_-] = 2J_3. \tag{9}$$

However, the hermiticity conditions for $SU(2)$ and $SU(1,1)$ are different in this basis, namely in any unitary representation we have

$$J_3 = J_3^+, \quad (J_+)^+ = J_- \quad \text{for } SU(2), \tag{10}$$

and

$$J_3 = J_3^+, \quad (J_+)^+ = -J_- \quad \text{for } SU(1,1). \tag{11}$$

The matrix elements of the generators in a canonical basis can be written in the same form[8]

$$J_\pm \,|\, JM\rangle = [(J \mp M)(J \pm M + 1)]^{1/2} \,|\, JM \pm 1\rangle,$$
$$J_3 \,|\, JM\rangle = M \,|\, JM\rangle \tag{12}$$

for both algebras and the Casimir operator in both cases satisfies

$$J^2 \,|\, JM\rangle = \{\tfrac{1}{2}(J_+ J_- + J_- J_+) + J_3^2\} \,|\, JM\rangle = J(J+1) \,|\, JM\rangle. \tag{13}$$

The numbers J and M, labeling representations and basis vectors, run through different values for $SU(2)$ and $SU(1,1)$, of course. Restricting ourselves to those irreducible unitary representations of the algebra that can be integrated to representations of the group, we find that[8]

$$J = 0, \tfrac{1}{2}, 1, \tfrac{3}{2}, \cdots, \quad M = -J, -J+1, \cdots, J \tag{14}$$

for $SU(2)$. For $SU(1,1)$ all the unitary representations, except the trivial one, are infinite-dimensional and we have several series of representations[8,9]:

(a) Discrete lower bounded series.

$$J = 0, \tfrac{1}{2}, 1, \tfrac{3}{2}, \cdots, \quad M = J+1, J+2, \cdots. \tag{15}$$

(b) Discrete upper bounded series.

$$J = 0, \tfrac{1}{2}, 1, \tfrac{3}{2}, \cdots, \quad M = -J-1, -J-2, \cdots. \tag{16}$$

(c) First continuous principal series.

$$J = -\tfrac{1}{2} + iq, \quad q = \text{real}, \quad M = 0, \pm 1, \pm 2, \cdots. \tag{17}$$

(d) Second continuous principal series.

$$J = -\tfrac{1}{2} + iq, \quad q = \text{real}, \quad M = \pm\tfrac{1}{2}, \pm\tfrac{3}{2}, \cdots. \tag{18}$$

(e) Continuous supplementary series.

$$-1 < J < 0, \quad M = 0, \pm 1, \pm 2, \cdots. \tag{19}$$

We shall normalize the $SU(1,1)$ states in the conventional manner, putting

$$\langle JM \,|\, J'M' \rangle = \delta(J - J')\delta_{MM'}, \tag{20}$$

where $\delta(J-J')$ is a Dirac δ function for continuous representations and a Kronecker delta for discrete ones.

3. GENERALIZED COUPLING FOR THE HIGHEST AND LOWEST $SU(2)$ STATES

In the following we shall denote $SU(2)$ multiplets as $|JM\rangle$ and $SU(1,1)$ ones as $|{}^J_M\rangle$. Let us now rewrite equation (1) for $M = J$ and derive an expression for the generalized coupling coefficient. We have

$$|JJ\rangle = \sum_r \begin{pmatrix} j & l & J \\ J+r & -r & J \end{pmatrix} \left| {}^{\,j}_{J+r} \right\rangle \left| {}^{\,l}_{-r} \right\rangle. \tag{21}$$

The range of summation in (21) depends on the type of $SU(1,1)$ representations figuring on the right-hand side of (21). The form of (21) ensures that $J_3 \,|\, JJ\rangle = J \,|\, JJ\rangle$ and immediately shows that the representations j and l must either both be continuous or both discrete, in which case one must be lower and one upper bounded. For definiteness we assume throughout that if j and l are discrete, then l is upper bounded, j lower bounded. If j and l are both continuous, then $-\infty < r < \infty$, if they are discrete, then $\max(l+1, j-J+1) \le r < \infty$. If l belongs to the first principal series, the supplementary series, or a discrete integer series, then r is integer, otherwise r is half-integer. If $J+r$ is integer, then j must correspond to the first principal series, the supplementary series or a discrete integer representation. If $J+r$ is half-integer, j must correspond to the second principal series or a discrete half-integer representation.

Let us now apply the raising operator J_+ to both sides of (21). We obtain

$$\begin{aligned} O = J_+ \,|\, JJ\rangle &= \sum_r \Bigg\{ \begin{pmatrix} j & l & J \\ J+r & -r & J \end{pmatrix} [(j-J+r)(j+J+r+1)]^{1/2} \\ &+ \begin{pmatrix} j & l & J \\ J+r+1 & -r-1 & J \end{pmatrix} \\ &\times [(l+r+1)(l-r)]^{1/2} \Bigg\} \left| {}^{\,j}_{J+r+1} \right\rangle \left| {}^{\,l}_{-r} \right\rangle, \end{aligned}$$

implying a simple recursion relation

$$\begin{pmatrix} j & l & J \\ J+r+1 & -r-1 & J \end{pmatrix} = -\left(\frac{(j-J-r)(j+J+r+1)}{(l+r+1)(l-r)} \right)^{1/2} \begin{pmatrix} j & l & J \\ J+r & -r & J \end{pmatrix}. \tag{22}$$

Relation (22) can be solved in such a manner as to express a general coefficient $\left({}^{\,j}_{J+r}\ {}^{\,l}_{-r}\ {}^{\,J}_{J}\right)$ in terms of some

"starting coefficient" which can be chosen to correspond to $r = r_{\min} = \max(l+1, j-J+1)$ if j and l are discrete, or to $r = 0$ ($r = \frac{1}{2}$) if both l and j belong to continuous series. It is easy to check that in all cases the solution of recursion relation (22) can be written as

$$\begin{pmatrix} j & l & J \\ J+r & -r & J \end{pmatrix} = (-1)^{r-l}\left(\frac{\Gamma(J-j+r)\Gamma(j+J+r+1)}{\Gamma(l+r+1)\Gamma(-l+r)}\right)^{1/2} x_0(j,l,J), \tag{23}$$

where x_0 is an arbitrary constant that can depend on the $SU(1,1)$ and $SU(2)$ representations that we are coupling (but not on r, i.e., the individual basis vector). The coefficient x_0 can be fixed, as we shall show in Sec. 6, by requiring that the generalized coupling coefficients should be the analytic continuation in j and l of $SU(2)$ Clebsch–Gordan coefficients. The expansion (21) can now be rewritten as

$$|JJ\rangle = x_0 \sum_r (-1)^{r-l} \times \left(\frac{\Gamma(J-j+r)\Gamma(j+J+r+1)}{\Gamma(l+r+1)\Gamma(-l+r)}\right)^{1/2} \left|\begin{matrix} j \\ J+r \end{matrix}\right\rangle \left|\begin{matrix} l \\ -r \end{matrix}\right\rangle, \tag{24}$$

where the sum ranges over all integer (half-integer) values of r for which the arguments of the Γ functions are not negative integers or zero.

Quite analogously we can put $M = -J$ in formula (1), apply J_- to both sides and obtain the generalized coupling coefficient for the lowest $SU(2)$ weight, namely

$$\begin{pmatrix} j & l & J \\ -J+r & -r & -J \end{pmatrix} = (-1)^{r-l}\left(\frac{\Gamma(r-l)\Gamma(r+l+1)}{\Gamma(r+j-J+1)\Gamma(r-j-J)}\right)^{1/2} \tilde{x}_0(j,l,J), \tag{25}$$

where $\tilde{x}_0(j,l,J)$ is again an unknown constant. We shall show below [see (29) and (30)] that

$$\tilde{x}_0 = x_0(-1)^J. \tag{26}$$

4. GENERALIZED COUPLING COEFFICIENTS FOR ARBITRARY STATES

Let us now apply the lowering operator J_- to Eq. (24) $(J-M)$ times. Using formulas (3) and performing some algebra, we find

$$\begin{aligned}(J_-)^{J-M}|JJ\rangle &= \left(\Gamma(2J+1)\frac{\Gamma(J-M+1)}{\Gamma(J+M+1)}\right)^{1/2}|JM\rangle \\ &= x_0\Gamma(J-M+1)(-1)^{(J-M)/2}\sum_r (-1)^{r-l} \\ &\times \frac{\Gamma(J+j+r+1)\Gamma(J-j+r)}{\Gamma(l+r+1)\Gamma(-l+r)} \\ &\times \sum_{p=r}^{J-M+r} \frac{1}{\Gamma(p-r+1)\Gamma(J-M-p+r+1)} \\ &\times \left(\frac{\Gamma(-l+p)\Gamma(l+p+1)}{\Gamma(j+M+p+1)\Gamma(-j+M+p)}\right)^{1/2} \\ &\times \left|\begin{matrix} j \\ M+p \end{matrix}\right\rangle \left|\begin{matrix} l \\ -p \end{matrix}\right\rangle,\end{aligned} \tag{27}$$

where both the r and p summation ranges are such that none of the arguments of the Γ functions are negative integers or zero.

Formula (27) can be further improved by interchanging the order of summation. The final result is

$$|JM\rangle = x_0\left(\frac{\Gamma(J+M+1)\Gamma(J-M+1)}{\Gamma(2J+1)}\right)^{1/2}(-1)^{(J-M)/2}\sum_{p=p_0}^{\infty}\left(\frac{\Gamma(-l+p)\Gamma(l+p+1)}{\Gamma(j+M+p+1)\Gamma(-j+M+p)}\right)^{1/2} \times \sum_{r=r_0}^{p}(-1)^{r-l}\frac{\Gamma(J+j+r+1)\Gamma(J-j+r)}{\Gamma(l+r+1)\Gamma(-l+r)\Gamma(p-r+1)\Gamma(J-M-p+r+1)}\left|\begin{matrix} j \\ M+p \end{matrix}\right\rangle \left|\begin{matrix} l \\ -p \end{matrix}\right\rangle, \tag{28}$$

where

$$p_0 = \begin{cases} \max(l+1, j-M+1) & \text{for } l \text{ and } j \text{ discrete} \\ -\infty & \text{for } l \text{ and } j \text{ continuous} \end{cases} \quad \text{and} \quad r_0 = \max(-J+M+p, l+1),$$

i.e., the summation limits are again determined by the Γ functions. Note that the r summation is always over a finite region.

Finally we must ensure that the general coupling coefficient

$$\begin{pmatrix} j & l & J \\ M+p & -p & M \end{pmatrix} = x_0(-1)^{(J-M)/2}\left(\frac{\Gamma(J+M+1)\Gamma(J-M+1)}{\Gamma(2J+1)}\frac{\Gamma(-l+p)\Gamma(l+p+1)}{\Gamma(j+M+p+1)\Gamma(-j+M+p)}\right)^{1/2} \times \sum_{r=r_0}^{p}(-1)^{r-l}\frac{\Gamma(J+j+r+1)\Gamma(J-j+r)}{\Gamma(l+r+1)\Gamma(-l+r)\Gamma(p-r+1)\Gamma(J-M-p+r+1)} \tag{29}$$

coincides for $M = -J$ with coefficient (25), calculated for the lowest $SU(2)$ weight (in other words we must check that $J^-|J-J\rangle = 0$). Comparing (25) and (29) we see that the two are consistent if and only if

$$F(J,j,l,p) \equiv \sum_{r=r_0}^{p}(-1)^{r-l}\frac{\Gamma(J+j+r+1)\Gamma(J-j+r)}{\Gamma(l+r+1)\Gamma(-l+r)\Gamma(p-r+1)\Gamma(2J-p+r+1)} = N(J,j,l)(-1)^{p-l-J} = (\tilde{x}_0/x_0)(-1)^{p-l-J}, \tag{30}$$

the essential point being that $N(J,j,l)$ does not depend on p.

We shall prove elsewhere[10] that

$$F(J,l+J-k,l,p)=(-1)^{p-l},\quad 0\le k\le 2J, \tag{31}$$

and that for $j\neq l+J-k$, $0\le k\le 2J$ the sum $F(J,j,l,p)$ cannot satisfy condition (30).

To summarize: Generalized $SU(2)$ multiplets, corresponding to angular momentum J, can be constructed from products of $SU(1,1)$ multiplets, corresponding to the representations j and l, providing that the numbers j, l and J satisfy the condition (2). The generalized coupling coefficients of Eq. (1) are then given by formula (29).

Note that condition (2) implies that either representations j and l are both discrete (then one must be upper and one lower bounded) or both continuous and then $j=l$. Note also that condition (2) is weaker than a triangular relation since it does not imply $J\le l+j$ (but does imply $|l-j|\le J$). In particular, if $l=j$ then J can be an arbitrary integer.

5. EXAMPLES

Not surprisingly, the general formula (29) for the coupling coefficients is quite complicated, not more so, however, than the corresponding formula, e.g., for $SU(2)$ Clebsch–Gordan coefficients. For low values of J the formulas become quite simple and we shall give several examples. Note that the formulas (23) and (25) for arbitrary J but $M=\pm J$ are also quite simple.

Let us spell out several special cases that can either be derived directly or obtained from (23), (25) and/or (29). We have

$J=0$.

$$\begin{pmatrix} l & l & 0\\ p & -p & 0\end{pmatrix}=x_0(-1)^{p-l}. \tag{32}$$

$J=\frac{1}{2}$.

$$\begin{pmatrix} j & l & \frac{1}{2}\\ \frac{1}{2}+p & -p & \frac{1}{2}\end{pmatrix}=\begin{cases}x_0(-1)^{p-l}(l+p+1)^{1/2}\\ x_0(-1)^{p-l}(-l+p)^{1/2}\end{cases}\quad \text{for } j=\begin{cases}l+\frac{1}{2}\\ l-\frac{1}{2}\end{cases} \tag{33}$$

$$\begin{pmatrix} j & l & \frac{1}{2}\\ -\frac{1}{2}+p & -p & -\frac{1}{2}\end{pmatrix}=\begin{cases}x_0(-1)^{p-l+1/2}(-l+p-1)^{1/2}\\ x_0(-1)^{p-l+1/2}(l+p)^{1/2}\end{cases}\quad \text{for } j=\begin{cases}l+\frac{1}{2}\\ l-\frac{1}{2}\end{cases}$$

$J=1$.

$$\begin{pmatrix} j & l & 1\\ 1+p & -p & 1\end{pmatrix}=\begin{cases}x_0(-1)^{p-l}[(-l+p+1)(-l+p)]^{1/2}\\ x_0(-1)^{p-l}[(-l+p)(l+p+1)]^{1/2}\\ x_0(-1)^{p-l}[(l+p+1)(l+p+2)]^{1/2}\end{cases}\quad \text{for } j=\begin{cases}l-1\\ l\\ l+1\end{cases}$$

$$\begin{pmatrix} j & l & 1\\ p & -p & 0\end{pmatrix}=\begin{cases}x_0(-1)^{p-l+1/2}[2(-l+p)(l+p)]^{1/2}\\ x_0(-1)^{p-l+1/2}\sqrt{2}p\\ x_0(-1)^{p-l+1/2}[2(l+p+1)(-l+p-1)]^{1/2}\end{cases}\quad \text{for } j=\begin{cases}l-1\\ l\\ l+1\end{cases} \tag{34}$$

$$\begin{pmatrix} j & l & 1\\ -1+p & -p & 1\end{pmatrix}=\begin{cases}x_0(-1)^{p-l+1}[(l+p)(l+p-1)]^{1/2}\\ x_0(-1)^{p-l+1}[(l+p)(-l+p-1)]^{1/2}\\ x_0(-1)^{p-l+1}[(-l+p-1)(-l+p-2)]^{1/2}\end{cases}\quad \text{for } j=\begin{cases}l-1\\ l\\ l+1\end{cases}$$

The general formula (29) also simplifies very significantly for the "stretched" coefficients, when $j=l+J$ or $j=l-J$. In these cases the summations can be performed and we obtain

$$\begin{pmatrix} l+J & l & J\\ M+p & -p & M\end{pmatrix}=x_0(-1)^{(J-M)/2+p-l}\left(\frac{\Gamma(2J+1)}{\Gamma(J-M+1)\Gamma(J+M+1)}\times\frac{\Gamma(J+M+l+p+1)\Gamma(-l+p)}{\Gamma(-J+M-l+p)\Gamma(l+p+1)}\right)^{1/2} \tag{35}$$

and

$$\begin{pmatrix} l-J & l & J\\ M+p & -p & M\end{pmatrix}=x_0(-1)^{(J-M)/2+p-l}\left(\frac{\Gamma(2J+1)}{\Gamma(J-M+1)\Gamma(J+M+1)}\times\frac{\Gamma(J+M+p-l)\Gamma(l+p+1)}{\Gamma(-l+p)\Gamma(l-J+M+p+1)}\right)^{1/2} \tag{36}$$

In order to illustrate the problem of normalizing the generalized $SU(2)$ multiplets [see formula (4)], let us consider a specific example, e.g., $J=0$ or 1, $M=0$, $j=l$:

$$|ll\,00\rangle=|00\rangle=x_0\sum_{p=l+1}^{\infty}(-1)^{p-l}\left|\begin{matrix}l\\-p\end{matrix}\right\rangle\left|\begin{matrix}l\\p\end{matrix}\right\rangle, \tag{37}$$

$$|ll\,10\rangle=|10\rangle=x_0\sqrt{2}\sum_{p=l+1}^{\infty}(-1)^{p-l+1/2}p\left|\begin{matrix}l\\-p\end{matrix}\right\rangle\left|\begin{matrix}l\\p\end{matrix}\right\rangle. \tag{38}$$

It is easy to see that the norms of these vectors are given by

$$\langle l'l'\,00|ll\,00\rangle=\delta_{ll'}|x_0|^2\sum_{p=l+1}^{\infty}p^0, \tag{39}$$

$$\langle l'l'\,10|ll\,10\rangle=\delta_{ll'^2}|x_0|^2\sum_{p=l+1}^{\infty}p^2, \tag{40}$$

and that both sums diverge (in general like p^{2J}).

It should be noted here that the states $|j\,l\,J\,M\rangle$ and $|j\,l\,J'\,M\rangle$ for $J\neq J'$ (but all other labels equal) are not orthogonal. For example,

$$\langle ll\,00|ll\,10\rangle=|x_0|^2\sqrt{2}\,i\sum_{p=l+1}^{\infty}p, \tag{41}$$

and the sum again diverges [although more slowly than the one in (40)]. In formulas (37)–(41) we assume for

definiteness that l is integer or half-integer. The lack of orthogonality for $J \neq J'$ shows clearly that in the space that we are considering the $SU(2)$ Casimir operator is at most a symmetric operator, but not a Hermitian one. In all our arguments the left-hand sides of the equations are defined by the right-hand sides; in particular, the $SU(1,1)$ states are normalized in a conventional manner. We did not investigate the problem of redefining the scalar product in the $SU(2)$ space in a more appropriate manner.

6. PROPERTIES OF THE GENERALIZED COUPLING COEFFICIENTS

A. Infinite norm

Using the formula (1) and the normalization (20) of $SU(1,1)$ states, we have

$$\langle J'M'|JM\rangle = \delta_{MM'}\delta(j-j')\delta(l-l')$$
$$\times \sum_p \begin{pmatrix} j & l & J \\ M+p & -p & M \end{pmatrix} \begin{pmatrix} j & l & J' \\ M+p & -p & M \end{pmatrix}^* . \quad (42)$$

One can show that the sum in (42) diverges rapidly, namely as $p^{J+J'}$ for $p \to \infty$. The general proof is somewhat cumbersome, so let us just illustrate this result by examples. For l being an upper-bounded series and $j = l \pm J$ we have $l+1 \leq p < \infty$ and using (35) or (36) we see that the general term in the series behaves as

$$\begin{pmatrix} l+J & l & J \\ M+p & -p & M \end{pmatrix} \begin{pmatrix} l+J' & l & J' \\ M+p & -p & M \end{pmatrix}^* \xrightarrow[p\to\infty]{}$$
$$|x_0|^2(-1)^{(J+J'-2M)/2} \left(\frac{\Gamma(2J+1)\Gamma(2J'+1)}{\Gamma(J-M+1)\Gamma(J+M+1)\Gamma(J'-M+1)\Gamma(J'+M+1)}\right)^{1/2} p^{J+J'} \quad (43)$$

(use was made of Stirling's asymptotic formula for Γ functions[11]). Similarly, for $J = J' = \pm M$, we have [see (23) or (25)]

$$\left|\begin{pmatrix} j & l & J \\ M+p & -p & M \end{pmatrix}\right|^2 \xrightarrow[p\to\infty]{} p^{2J}. \quad (44)$$

B. Relation to $SU(2)$ and $SU(1,1)$ Clebsch-Gordan coefficients

The method used in Secs. 3 and 4 to derive expressions for the generalized coupling constants can be applied directly to derive analogous formulas for $SU(2)$ or $SU(1,1)$ coupling coefficients. This is particularly simple in the case of $SU(2)$ multiplets. Treating the multiplets on the right-hand side of (21) as $SU(2)$ multiplets, going through the procedure of Sec. 3, making use of the usual normalization

$$\langle J'M'|JM\rangle = \delta_{JJ'}\delta_{MM'} \quad (45)$$

of ordinary $SU(2)$ multiplets, and making the usual phase conventions,[6] we obtain an expression for a "highest weight" $SU(2)$ Clebsch-Gordan coefficient:

$$(j\,J\,-m\;l\,m\,|JJ) = (-1)^{j-J+m}$$
$$\times\left(\frac{\Gamma(2J+2)\Gamma(j+l-J+1)}{\Gamma(j+l+J+2)\Gamma(j-l+J+1)\Gamma(-j+l+J+1)}\right.$$
$$\left.\times\frac{\Gamma(j+J-m+1)\Gamma(l+m+1)}{\Gamma(j-J+m+1)\Gamma(l-m+1)}\right)^{1/2}. \quad (46)$$

Formula (46) should be compared with (23) for the highest weight generalized coupling coefficients. Making use of standard properties of Γ functions, in particular $\Gamma(z)\Gamma(1-z) = \pi/\sin\pi z$, we find that the two formulas coincide, if we define x_0 to be

$$x_0 \equiv x_0(j,l,J) = e^{(i\pi/2)(j-J-l+2)}$$
$$\times\left(\frac{\Gamma(j+l-J+1)\Gamma(2J+2)}{\Gamma(j+l+J+2)\Gamma(j-l+J+1)\Gamma(-j+l+J+1)}\right)^{1/2}. \quad (47)$$

Proceeding along the lines of Sec. 4, i.e., applying the operator $(J_-)^{J-M}$ to the state $|JJ\rangle$ we derive a general expression for the $SU(2)$ Clebsch-Gordan coefficient. The formula coincides with expression 13.1b of Yutsis and Bandzaitis[6] and can be shown to agree with our formula (29) if x_0 is given by (47).

Thus our generalized coupling coefficients (29) can be considered to be an analytic continuation of the usual $SU(2)$ Clebsch-Gordan coefficients. The continuation is in the quantum number p, from $-l \leq p \leq l$ to $p \geq l+1$ for discrete representations of $SU(1,1)$, or to arbitrary integer or half-integer $-\infty < p < \infty$ for continuous representations. The continuation is also in j and l — from integer or half-integer values, satisfying a triangular relation with J, to arbitrary j and l, satisfying (2).

Since the $SU(1,1)$ Clebsch-Gordan coefficients[7] have been shown to be analytic continuations of the $SU(2)$ ones, our coefficients (29) can equally well be considered to be analytic continuations in J and M of the $SU(1,1)$ Clebsch-Gordan coefficients.

C. Further comments

When considering Clebsch-Gordan coefficients for any group, properties that are usually stressed are symmetries, recursion relations, and the relation between the coefficients and D functions (finite transformation matrices). This last question cannot even be considered here, since we have not established whether the constructed representation of the $SU(2)$ algebra can be integrated to a representation of the group. Hence we do not know whether the D functions exist at all, still less what they are. The symmetry properties of the generalized coupling coefficients are not particularly illuminating. Consider for instance an interchange of j and l. The corresponding $SU(1,1)$ representations can both be discrete—one upper bounded, one lower bounded—then their interchange is not meaningful. Alternatively, both representations can be continuous, but then we must have $j = l$ and the interchange is trivial. Recursion relations can readily be derived by applying the raising and lower-

ing operators $J_\pm$ to both sides of Eq. (1). We shall not present the resulting formulas here.

D. Other types of coupling coefficients

Finally, let us mention that other couplings involving $SU(2)$ and $SU(1,1)$ multiplets can be considered. Thus we can write

$$\left|\begin{matrix} l \\ -p \end{matrix}\right\rangle = \sum_{M=-J}^{J} \left\{\begin{matrix} j & J & l \\ -p-M & M & -p \end{matrix}\right\} \left|\begin{matrix} j \\ -p-M \end{matrix}\right\rangle |JM\rangle, \tag{48}$$

where we are constructing an $SU(1,1)$ multiplet from products of $SU(1,1)$ and $SU(2)$ multiplets. Proceeding similarly as above we can show that the coefficients $\left\{\begin{matrix} j & J & l \\ -p-M & M & -p \end{matrix}\right\}$ in (48) exist, providing that j, l and J satisfy (2). The representation l and j must both be upper bounded or both lower bounded, if j and l are discrete. It is quite easy to construct the coefficients in (48) explicitly and also to relate them to the coefficients (29), but we shall not do this here.

The remaining two types of conceivable couplings, namely $SU(2) \times SU(2) \to SU(1,1)$ and $SU(1,1) \times SU(2) \to SU(2)$ are actually not possible. Consider the coupling of two $SU(2)$ states to an $SU(1,1)$ state, e.g.,

$$\left|\begin{matrix} l \\ -p \end{matrix}\right\rangle = \sum_{M=-J}^{J} \left[\begin{matrix} L & J & l \\ -p-M & M & -p \end{matrix}\right] |L \quad -p-M\rangle\, |JM\rangle, \qquad p \geq l+1. \tag{49}$$

If the left-hand side is supposed to be an upper bounded $SU(1,1)$ multiplet, then we must have

$$(J_-)^k \left|\begin{matrix} l \\ -p \end{matrix}\right\rangle \neq 0 \quad \text{for all } k. \tag{50}$$

On the right-hand side we have

$$(J_-)^k\, |L \quad -p-M\rangle\, |JM\rangle = \sum_{s=0}^{k} \binom{k}{s} (J_-)^{k-s}\, |L \quad -p-M\rangle (J_-)^s |JM\rangle. \tag{51}$$

However,

$$(J_-)^{k-s}\, |L \quad -p-M\rangle = 0 \quad \text{for } -p-M-k+s < -L,$$
$$(J_-)^s |JM\rangle = 0 \quad \text{for } M-s < -J,$$

so that for

$$k > L+J-p$$

all terms in (51) are zero and hence the multiplet defined by (49) has a finite number of components and cannot transform under a unitary representation of $SU(1,1)$.

Similarly, we cannot construct an $SU(2)$ multiplet out of the product of an $SU(2)$ and an $SU(1,1)$ multiplet. Indeed, put

$$|JM\rangle = \sum_{m=-L}^{L} \left[\begin{matrix} j & L & J \\ M-m & m & M \end{matrix}\right] \left|\begin{matrix} j \\ M-m \end{matrix}\right\rangle |Lm\rangle. \tag{52}$$

The condition for $|JM\rangle$ to be an $SU(2)$ multiplet is

$$(J_-)^{J+M+1} |JM\rangle = 0, \qquad (J_+)^{J-M+1} |JM\rangle = 0. \tag{53}$$

Applying both operators to the right-hand side of (52), we find that (53) imposes incompatible conditions.

7. CONCLUSIONS

Let us add a few more words about the possible significance and applications of the generalized coupling coefficients introduced in this paper. The connection with tachyons was mentioned in the Introduction. It is well-known[12] that the direct product of two tachyon-like representations of the Poincaré group (with $m_1^2 < 0$ and $m_2^2 < 0$) contains also tardyon-like representations (with $m^2 > 0$). It is hence perfectly meaningful to consider vertices, involving two tachyons and a tardyon. The corresponding Clebsch–Gordan coefficients of the Poincaré group have been calculated,[12] using the canonical linear momentum basis,[1] in which the four components of the linear momentum P_μ and one component of the spin operator $W_\mu = \epsilon_{\mu\nu\lambda\rho} M_{\nu\lambda} p_\rho$ is diagonal.

Alternatively, one can consider representations of the Poincaré group in an angular momentum basis. For positive mass squared $m^2 > 0$ representations such a basis[13] corresponds to diagonalizing the energy P_0, the square of the angular momentum J^2, a component J_3 and the helicity operator $\frac{1}{2}W_0 = \mathbf{PJ}$. A similar "angular momentum" basis can be introduced for tachyons by diagonalizing a component of the momentum P_3, the $SU(1,1)$ "angular momentum" $J_3^2 - K_1^2 - K_2^2$ and "helicity" $-\frac{1}{2}W_3 = M_{12}p_0 + M_{01}p_2 + M_{20}p_1$ and the component J_3. If we calculate the Clebsch–Gordan coefficients of the Poincaré group in such a basis, then the coupling coefficients of this article will make an appearance. The selection rules (2) and those of Sec. 3 can thus be interpreted as angular momentum selection rules, e.g., for tachyon pair production. We plan to return to these problems in a separate study.

A mathematical by-product of the approach of this paper is the derivation of combinatorial identities of the type illustrated by formulas (30) and (31). A derivation of this nontrivial summation formula, together with a description of Lie algebra methods of deriving more general identities of this type will be presented separately.[10]

Finally, let us mention that the approach of this paper can be applied for other Lie algebras. We have actually mainly made use of the fact that the $SU(2)$ and $SU(1,1)$ algebras have the same complex extension $SL(2,C)$, which enabled us to write the two algebras in the same form (see Sec. 2). Thus it should be possible to construct, e.g., $SU(n)$ [or $O(n)$] multiplets out of products of $SU(n-q,q)$ [or $O(n-q,q)$] multiplets.

ACKNOWLEDGMENTS

In conclusion we thank Dr. A. K. Bose, Dr. E. G. Kalnins, and Dr. R. T. Sharp for many helpful discussions. One of the authors (J.P.) takes the opportunity to thank the Aspen Center for Physics where part of this investigation was performed.

[1]E. P. Wigner. Ann. Math. **40**, 149 (1939).
[2]O. M. P. Bilaniuk, V. K. Deshpande, and E. C. G. Sudarshan. Am. J. Phys. **30**, 718 (1962). G. Feinberg. Phys. Rev. **159**, 1089 (1967). E. C. G. Sudarshan. Ark. Fys. **39**, 585 (1969).
[3]M. Toller, Nuovo Cim. **54A**, 295 (1968). P. D. B. Collins and E. J. Squires, *Regge Poles in Particle Physics* (Springer, Berlin, 1968). J. F. Boyce. J. Math. Phys. **8**, 675 (1967). P. Winternitz, "Expansions of Relativistic Scattering Amplitudes and Harmonic Analysis on the Lorentz Group" in *1970-1971 Lecture Series,* edited by D. E. Myers (The University of Arizona, Dept. of Math., Tucson, 1972).
[4]L. Armstrong, Jr., Phys. Rev. **A3**, 1546 (1971); L. Armstrong, Jr., J. Math. Phys. **12**, 953 (1971); M. J. Cunningham, "Lecture at Rutherford Centennial Symposium" in *The Structure of Matter* (University of Canterbury, New Zealand, 1972).
[5]P. Roman, J. J. Aghassi, and P. L. Huddleston. J. Math. Phys. **13**, 1852 (1972). L. Clavelli and P. Ramond. Phys. Rev. D **2**, 973 (1970). C. S. Kalman and J. Patera. Lett. Nuovo Cim. **5**, 78 (1972).
[6]E. P. Wigner, *Group Theory and Its Applications to the Quantum Mechanics of Atomic Spectra* (Academic Press, New York, 1959). A. P. Yutsis and A. A. Bandzaitis, *Theory of Angular Momentum in Quantum Mechanics* (Mintis, Vilnius, 1965) (in Russian).
[7]W. J. Holman and L. C. Biedenharn, Ann. Phys. **39**, 1 (1966); Ann. Phys. **47**, 205 (1968). L. Pukanszky. Amer. Math. Soc. Transl. **100**, 116 (1961). S. S. Sannikov. Dokl. Ak. Nauk SSSR **171**, 1058 (1966) [Sov. Phys. Dokledy **11**, 1045 (1967)].
[8]N. Ya. Vilenkin, *Special Functions and the Theory of Group Representations* (Am. Math. Soc., Providence, R.I. 1968).
[9]V. Bargmann. Ann. Math. **59**, 1 (1954). I. M. Gel'fand, M. I. Graev, and N. Ya Vilenkin, *Generalized Functions. Integral Geometry and Representation Theory* (Academic Press, New York, 1966), Vol. 5.
[10]J. Patera and P. Winternitz, "Lie Algebras and Combinatorial Identities" (to be published).
[11]A. Erdelyi *et al, Higher Transcendental Functions* (McGraw-Hill, New York, 1953), Vol. 1.
[12]M. Schaaf, *The Reduction of the Product of Two Irreducible Unitary Representations of the Proper Ortochronous Quantum Mechanical Poincaré Group* (Springer, Berlin, 1970).
[13]J. S. Lomont and H. E. Moses. J. Math. Phys. **5**, 294, 1438 (1964).

On the transport properties of a van der Waals fluid. I. Formal theory*

J. Piasecki and P. Resibois

Faculté des Sciences, Université Libre de Bruxelles, Belgium
(Received 6 March 1972)

We present a formal exact analysis of the autocorrelation formulas for transport coefficients in a van der Waals fluid. We show that the first correction due to the long range force is of order γ (γ = inverse range of the force) and the formal expression for this correction is displayed in terms of the Prigogine–Balescu graphs. The analysis crucially rests upon a generalization of hydrodynamical modes in a van der Waals fluid, valid for wave numbers smaller or of the order of γ.

I. INTRODUCTION

It is now almost a century ago that van der Waals had the remarkable idea of analyzing the equilibrium properties of a classical fluid by separating the pair interaction between the particles into a short-range strongly repulsive part and a long-range part, corresponding to the weak attraction.

Although the qualitative success of this theory was rapidly recognized, it is only recently that a rigorous formulation of this theory has been given, both for the one-dimensional case[1] and in three dimensions.[2,3] Appropriate perturbation techniques have also been developed.[4–6] The starting point is to write the potential of the pair interaction in the form (we work in three dimensions):

$$V(r) = V^s(r) + \gamma^3 V^L(\gamma r), \tag{I.1}$$

where $V^s(r)$ is the short-range part while $\gamma^3 V^L(\gamma r)$ (the so-called "Kac potential," originally introduced by Brout[7] in the similar problem of ferromagnetism) is the long-range part of the total potential. The parameter γ measures the inverse range of V^L and the γ-independent integral $\gamma^3 \int d^3r V(\gamma r)$ is supposed to converge; γ thus plays the role of a smallness parameter and various perturbation methods have been developed in order to obtain γ expansions for the thermodynamic properties and the correlation functions of the system. In these expansions, the role of the reference system is played by a fluid of particles interacting via the short-range potential only.

Under certain assumptions about the functions V^s and V^L (see Refs. 3 and 5), it has been proved that the $\gamma \to 0$ limit leads to the van der Waals type equation of state combined with the Maxwell equal area construction. The lowest order corrections to this equation as well as to the thermodynamic and correlation functions have been calculated.

Apart from some semimacroscopic studies close to the critical region (where precisely, for γ finite, the γ expansions are known to fail!),[8] there has been no similar development for the transport properties of a van der Waals fluid. The purpose of the present paper is to present such a theory: We want to determine the lowest order correction to the transport coefficients of a fluid resulting from the presence of the long-range potential $\gamma^3 V^L(\gamma r)$, when this fluid is far from its critical region.

As is often the case in many-body theory, the feasibility of this program depends fairly heavily upon a suitable formal technique. This technique is presented in Sec. II, where we start our analysis from the well-known Green–Kubo expressions for transport coefficients. We then recall briefly the many-body analysis of these formulas according to the method of Prigogine and coworkers[9–11]; in particular, we use their diagram technique and we show that, due to the linear nature of the problem, the technique recently applied by one of us (J. P.)[12] to the study of linearized kinetic equations can be transposed here to considerably simplify the formal exact expressions for these transport coefficients. The developments given in this section are valid irrespective of the decomposition (I. 1) and, as a matter of fact, they only reproduce previously published calculations in a manner which is convenient for our present purpose—we will thus be rather brief and refer the reader to the literature for further details.

In Sec. III, we analyze the γ expansion of one of the central operators of Sec. II, namely the so-called collision operator which is the generalization for dense fluids of the well-known Boltzmann operator of dilute gases. We show that a naive expansion of this operator in power of V^L fails to converge for γ small and that, in order to get convergent expressions, we first have to replace the free propagator of the particles by a "dressed propagator" which, in the long wavelength limit, describes the hydrodynamical propagation of these particles. The physical picture underlying this renormalization procedure is quite simple and can be interpreted in very intuitive terms: Because the Kac potential is very long range, the particles will feel it in their motion over long distances and long times. For such extended time and space intervals however, it is unrealistic to approximate the dynamics of these particles by their free particle motion; because the fluid in which they propagate is dense, we have rather to consider their hydrodynamic behavior in this fluid. Once this renormalization has been made, it is a straightforward, albeit nontrivial, matter to establish the convergence of the γ expansion; more precisely, we give the rules which allow one to establish a lower bound to the γ dependence of any graph that contributes to the collision operator. We show in particular that the dominant correction due to the Kac potential is of order γ^1. This result is in contrast with the corresponding expansion for the equilibrium properties, which generally involve a γ-independent correction. In Sec. IV, we extend the results of Sec. III to the other operators that are involved in the microscopic expression for the transport coefficients. As the calculations are similar to those presented in detail for the collision operator, we do not give here any detailed proof and merely present the results. In particular, the leading correction to any transport coefficient is again of order γ^1 and its

formal microscopic expression is presented graphically.

The results of Secs. III and IV are obtained under the assumption that the hydrodynamical modes of the van der Waals fluid can be expressed, to lowest order in γ, in terms of γ^0-order quantities. This assumption is consistently checked in Sec. V, using a slight generalization of the microscopic theory of hydrodynamical modes recently developed by one of the authors (P. R.).[13] In particular, it is shown there that the eigenvalues of these modes, denoted by $\Lambda^k_\alpha(\gamma)[\alpha \in (1,2\ldots5)]$, are given, *to lowest order in* γ, by

$$\Lambda^k_{1,2} = \pm ic(k\gamma^{-1})k - \Gamma(k\gamma^{-1})k^2, \quad \Lambda^k_{3,4} = -\eta^s k^2/n,$$
$$\Lambda^k_5 = -\kappa^s k^2/nC_p(k\gamma^{-1}), \tag{I.2}$$

where

$$\Gamma(k\gamma^{-1}) = \frac{1}{2n}\left[\frac{4}{3}\eta^s + \xi^s + \left(\frac{1}{C_v(k\gamma^{-1})} - \frac{1}{C_p(k\gamma^{-1})}\right)\kappa^s\right]. \tag{I.3}$$

Here η^s, ξ^s, and κ^s, respectively, denote the coefficients of shear viscosity, bulk viscosity, and thermal conductivity of the reference *hard-core* fluid, while $c(k\gamma^{-1})$, $C_p(k\gamma^{-1})$, and $C_v(k\gamma^{-1})$, respectively, are the sound velocity and specific heats of the van der Waals gas, calculated to zeroth order in γ and suitably generalized to small but nonzero wave number k ($k \lesssim \gamma$).

Let us stress that the present paper is only concerned with *formal* results, which, although exact (in the sense of formal perturbation theory), cannot be directly used to explicitly compute the first order correction to transport coefficients of a van der Waals gas. This explicit calculation can, however, be pursued up to the point where the aforementioned corrections are reduced to simple quadratures: This very important aspect of the present theory will be dealt with in a forthcoming publication.

Finally, let us close this introduction by one remark concerning the approach to equilibrium of a van der Waals fluid: Indeed, another important problem—independent of the calculation of transport coefficients—is the time behavior of the system when it is put in an arbitrary nonequilibrium state. Unfortunately, it seems that the present approach cannot be directly applied to the study of this question and that the γ expansion of the generalized time dependent kinetic equation requires a separate analysis. This analysis should presumably take careful account of the existence of two time scales in the system, related to the short- and long-range interactions, respectively. We hope to discuss this problem in a separate publication. Some mathematical details have also been relegated in appendixes.

II. FORMAL PRELIMINARIES

We consider a classical N-particle system with coordinates $r = r_1, r_2 \cdots r_N$ and $v = v_1, v_2 \cdots v_N$.[14] Its Hamiltonian is

$$H = \sum_{a=1}^{N} \frac{v_a^2}{2} + \sum_{a>b=1}^{N} V(|r_a - r_b|) \tag{II.1}$$

and the corresponding Liouville operator L is written

$$L = L_0 + \delta L, \tag{II.2}$$

where

$$L_0 = -i \sum_{a=1}^{N} v_a \frac{\partial}{\partial r_a}, \tag{II.3}$$

$$\delta L = \sum_{b>a=1}^{N} \delta L^{ab} = i \sum_{b>a=1}^{N} \frac{\partial V(rab)}{\partial r_a}\left(\frac{\partial}{\partial v_a} - \frac{\partial}{\partial v_b}\right). \tag{II.4}$$

We want to evaluate a transport coefficient X through the Green–Kubo correlation function:

$$X = \lim_{t\to\infty}\lim_{\infty} \frac{\beta}{\Omega}\int_0^t d\tau \int dr dv\, J^X \exp(-iL\tau)(J^X - \delta J^X)\rho^{eq}. \tag{II.5}$$

Here $\lim_\infty$ denotes the thermodynamic limit; Ω is the volume of the system. Moreover, $J^X = \sum_{i=1}^N J_i^X$ is the flow operator associated with X and δJ^X denotes the "counterterm" which insures that $(J^X - \delta J^X)$ is orthogonal to the invariants of L; for example, if we consider the thermal conductivity $X \equiv T\kappa$, we have

$$J^{T\kappa} = \sum_{i=1}^{N} J_i^{T\kappa}, \quad \delta J^{T\kappa} = \sum_{i=1}^{N} \delta J_i^{T\kappa}, \tag{II.6}$$

$$J_i^{T\kappa} = v_{ix}\left(\frac{v_i^2}{2} + \frac{1}{2}\sum_{j\neq i} V(r_{ij})\right) - \frac{1}{2}\sum_{j\neq i} v_i \frac{\partial V}{\partial r_{ij}}(r_{ij})_x, \tag{II.7}$$

$$\delta J_i^{T\kappa} = (h/n)v_{ix}, \tag{II.8}$$

where h is the equilibrium enthalpy density. Finally, the equilibrium distribution for N particles at temperature $T = 1/k\beta$ has been denoted by ρ^{eq}.

As was shown for example in Ref. 11, Eq. (II.5) can be transformed by introducing the four operators $\Psi_0(z)$ (collision operator), $\mathscr{C}_{k0}(z)$ (creation operator), $\mathscr{D}_{0k}(z)$ (destruction operator), and $\mathscr{P}_{kk'}(z)$ (propagation operator) which play a fundamental role in the kinetic theory of homogeneous systems developed by Prigogine and coworkers.[9,10] If we introduce the abbreviation $(k|A|k')$ for the Fourier transform,

$$\Omega^{-N}\int dr \exp\left(-i\sum_{a=1}^{N} k_a r_a\right) A(r,v) \exp\left(i\sum_{b=1}^{N} k'_b r_b\right), \tag{II.9}$$

the definitions of these operators are

$$\Psi_0(z) = \sum_{n=1}^{\infty}\left(0\left|-\delta L\left(\frac{1}{L_0 - z}\,Q(-\delta L)\right)^n\right|0\right), \tag{II.10}$$

$$\mathscr{C}_{k0}(z) = \sum_{n=1}^{\infty}\left(k\left|\left(\frac{1}{L_0 - z}\,Q(-\delta L)\right)^n\right|0\right), \tag{II.11}$$

$$\mathscr{D}_{0k}(z) = \sum_{n=1}^{\infty}\left(0\left|\left((-\delta L)Q\frac{1}{L_0 - z}\right)^n\right|k\right), \tag{II.12}$$

and

$$\mathscr{P}_{kk'}(z) = \sum_{n=0}^{\infty}\left(k\left|\frac{1}{L_0 - z}\left(Q(-\delta L)\frac{1}{L_0 - z}\right)^n\right|k'\right), \tag{II.13}$$

where $Q = I - |0)(0|$ (I is the identity operator), and z is a complex variable lying in the upper half-plane. When z is in the lower half-plane, we should take the analytical continuation of these expressions. Moreover, k and k' denote sequences $(k_1, \ldots, k_N)$ and $(k'_1, \ldots, k'_N)$, respectively, which both contain some nonzero wave vectors. (For a finite system, owing to the periodicity conditions, wave vector k_i has the form $2\pi n/\Omega^{1/3}$, where n is a vector with integer components. See Refs. 9 and 10).

In terms of these operators, the autocorrelation function (II. 5) can be rewritten in the form [see Eqs. (3. 12), (3. 13), (3. 13′) in Ref. 11]

$$X = X' + X'', \tag{II. 14}$$

where

$$X' = \lim_{t\to\infty}\lim_{\infty} \frac{i\beta}{2\pi\Omega}\int_0^t d\tau \oint_c \exp(-iz\tau)dz \int dv \times\left[(0|J^X|0) + \sum_{k\neq 0}(0|J^X|k)\mathscr{C}_{k0}(z)\right]\frac{1}{-z-\Psi_0(z)} \times\left[(0|(J^X-\delta J^X)\rho^{eq}\Omega^N|0) + \sum_{k'\neq 0}\mathscr{D}_{0k'}(z)\right] \times(k'|(J^X-\delta J^X)\rho^{eq}\Omega^N|0) \tag{II. 15}$$

and

$$X'' = \lim_{t\to\infty}\lim_{\infty} \frac{i\beta}{2\pi\Omega}\int_0^t d\tau \oint_c \exp(-iz\tau)dz \int dv \times\left[\sum_{k,k'\neq 0}(0|J^X|k)\mathscr{P}_{kk'}(z)(k'|(J^X-\delta J^X)\rho^{eq}\Omega^N|0)\right]. \tag{II. 16}$$

In these equations, the path C of complex integration lies in the upper half-plane and is parallel to the real axis.

At this point, it is worthwhile to make the following remarks:

(1) In Eq. (II. 5), the evolution operator $\exp(-iL\tau)$ acts upon the "initial condition" $(J^X - \delta J^X)\rho^{eq}$ which can be looked upon as a *linearized* deviation from the absolute equilibrium distribution ρ^{eq}. For instance, in the case of thermal conductivity, we have

$$(0|(J^{T\kappa} - \delta J^{T\kappa})\rho^{eq}\Omega^N|0) = \sum_{a=1}^{N}\delta\varphi(v_a)\prod_{b\neq a}\varphi^{eq}(v_b), \tag{II. 17}$$

where $\varphi^{eq}(v_1)$ denotes the equilibrium Maxwellian, while

$$\delta\varphi(v_a) = v_{ax}\left(\frac{v_a^2}{2} + nV_0 - \frac{h}{n}\right)\varphi^{eq}(v_a). \tag{II. 18}$$

Hence, taking the velocity integration

$$\int dv \equiv \int dv_1 \cdots dv_N$$

into account, we can further simplify expressions (II. 15) (II. 16) for X' and X'' by replacing the many-particle operators (II. 10)—(II. 13) by the corresponding *linearized* operators $\Psi_0^l(z)$, $\mathscr{C}_{k0}^l(z)$, $\mathscr{D}_{0k}^l(z)$ and $\mathscr{P}_{kk'}^l(z)$ (to be defined below), which only depend on a single velocity variable. This linearization procedure has been discussed in detail in Ref. 12 and the results concerning structure of the linearized operators obtained there can be directly applied to the present case; it will be discussed in more detail in the following.

(2) Similarly, the complex z integral and the time integral could have been directly performed in Eqs. (II. 15), (II. 16), leaving us merely with the bracketed contributions taken at

$$z = +i\epsilon.$$

However, we have not performed these integrations immediately in order to stress the similarity of the present equations with the formulas discussed in Ref. 12.

In order to make clear the simplifications which are brought into the problem by the linear nature of our "initial condition" $(J^X - \delta J^X)\rho^{eq}$, it is most convenient to use the diagrammatic technique of Prigogine and Balescu.[9,10] In order to avoid undue repetition, we shall, however, not give in detail here the rules which establish a one-to-one correspondence between graphs and analytical contributions and we urge the reader interested in the technical aspects of our discussion to constantly refer to Refs. 9 and 10. Nevertheless, we shall illustrate the main features of this graphical analysis by simple examples which should be understandable without serious difficulties.

Representing the elementary vertices, which are explicitly given by [see (II. 4), (II. 9)]

$$(k|\delta L^{ij}|k') = \Omega^{-1}V_{|k_i-k_i'|}(k_i'-k_i)\left(\frac{\partial}{\partial v_i} - \frac{\partial}{\partial v_j}\right) \times\delta^{K\tau}_{k_i'+k_j',\,k_i+k_j}\prod_{a\neq i,j}\delta^{K\tau}_{k_a,k_a'}, \tag{II. 21}$$

by the graphs of Figs. 1, we can, for example, describe the collision operator $\Psi_0(z)$ [see (II-10)] by the general structure of Fig. 2(a): Reading the graph from right to left, we start with no line (corresponding to a state $k=0$) and we combine the vertices of Fig. 1 in such a way that no intermediate state has no line [corresponding to the projector Q in Eq. (II-10)] until we arrive at the final state which again corresponds to no line ($k=0$).

If we now analyze this general structure in a slightly more detailed way, we notice that it can be realized in two different ways, as is illustrated in Fig. 2(b) and (c):

(1) *By semiconnected graphs*[9,12] In these there are two (or more) disconnected parts involving groups of particles which only have one particle (s) in common; by the law of conservation of wave vectors $\Sigma_a k_a = 0$, it is easily seen that, in some intermediate states, particle s has zero wave number, $k_s = 0$. In order to keep track of it, we then represent it by a dotted line (this graphical rule will be generally adopted in similar circumstances).

(2) *By fully-connected graphs*[9,11] These do not have the above property.

For the other operators $\mathscr{D}$, $\mathscr{C}$, $\mathscr{P}$, we can similarly define fully-connected and semiconnected contributions. It is the central result obtained in Ref. 12 that in the study of a *linearized* kinetic equation, the various semiconnected contributions exactly cancel each other, as a consequence of the equation $L\rho^{eq} = 0$. Exactly the same property holds here [see in particular the analogy between our Eq. (II. 15) and the starting equation (I. 3) of Ref. 12] and we thus merely state the result:

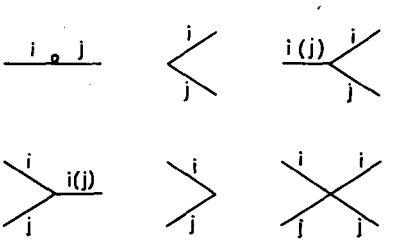


FIG. 1. The elementary vertices.

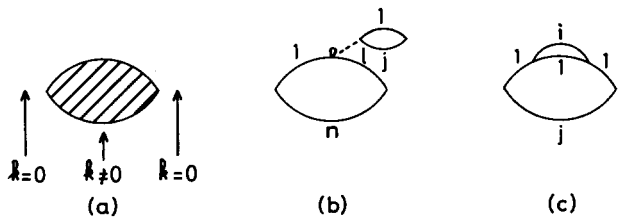


FIG. 2. The collision operator $\Psi_0(z)$: (a) general structure (b) semiconnected graph (c) fully connected graph.

$$X' = \lim_{t\to\infty} \frac{i\beta n}{2\pi} \int_0^t d\tau \oint_c \exp(-iz\tau)dz \int dv_1 \Big[(0|J_1^x|0) + \sum_{k\neq 0} (0|J_1^x|k)\mathscr{C}_{k0}^1(v_1;z)\Big]\Big(\frac{1}{-z-\Psi_0^1(v_1;z)}\Big) \times\Big\{\Big[\int dv^{N-1}(0|(J^x-\delta J^x)\rho^{eq}\Omega^N|0)\Big] + \mathscr{D}_0^1(v_1;z)\Big\} \tag{II.22}$$

and

$$X'' = \lim_{t\to\infty} \frac{i\beta n}{2\pi} \int_0^t d\tau \oint_c \exp(-iz\tau)dz \int dv_1 \sum_{k\neq 0} (0|J_1^x|k) \times \mathscr{P}_k^1(v_1;z). \tag{II.23}$$

In these equations, $\mathscr{C}_{k0}^1$ and Ψ_0^1 are one-particle operators while $\mathscr{P}_k^1$ and $\mathscr{D}_0^1$ are one-particle functions. More precisely, when $\Psi_0^1(v_1;z)$ acts on an arbitrary function of v_1 [denoted by $\Phi(v_1)$], we have[15]

$$\Psi_0^1(v_1;z)\Phi(v_1) = \lim_{\infty} \sum_{n=1}^{\infty}\sum_{s=1}^{N} \int dv^{N-1} \Big\{\Big(0\Big|_1\Big[-\delta L \times\Big(\frac{1}{L_0-z}Q(-\delta L)\Big)^n\Big]_s^{F.C.}\Big|0\Big)\Phi(v_s) \times \prod_{j\neq s}\varphi^{eq}(v_j) + \sum_{\substack{k\neq 0\\ k_s=0}} \Big(0\Big|_1\Big[-\delta L\Big(\frac{1}{L_0-z} \times Q(-\delta L)\Big)^n\Big]_s^{F.C.}\Big|k\Big)(k|\rho^{eq}\Omega^N|0) \times\Phi(v_s)/\varphi^{eq}(v_s)\Big\}. \tag{II.24}$$

In the notation ${}_i[\ ]_j^{F.C.}$, the superscript F.C. indicates that only fully-connected contributions should be retained while the subscripts i and j on the left and on the right of the brackets mean that the vertex at the extreme left and the vertex on the extreme right should, respectively, involve particle i and j [see the analog equation (3.16) of Ref. 12].

Similarly, we have

$$\mathscr{D}_0^1(v_1;z) = \lim_{\infty} \sum_{n=1}^{\infty}\sum_{s=1}^{N} \int dv^{N-1} \sum_{\substack{k\neq 0\\ k_s\neq 0}} \times\Big(0\Big|_1\Big[\Big(-\delta L Q\frac{1}{L_0-z}\Big)^n\Big]_s^{F.C.}\ k\Big) \times(k|(J_s^x-\delta J_s^x)\rho^{eq}\Omega^N|0). \tag{II.25}$$

Similar formulas hold for $\mathscr{C}_{k0}^1(v_1;z)$ and $\mathscr{P}_k^1(v_1;z)$; they are obtained by replacing the ket $(0|$ by $(k|(L_0-z)^{-1}$, respectively, at the left of Eqs. (II.24) and (II.25).

Note that although the thermodynamic limit is taken in these formulas, we find it convenient to formally keep discrete sums over the wave vectors k: They should, in explicit calculations, he replaced by appropriate integrals. Let us also stress that the transformations leading from (II.15), (II.16) to (II.22), (II.23) are *exact*; indeed, the linear nature of the response function (II.5) is not an *ad hoc* assumption but is really built into this equation, as is examplified in (II.17).

An important feature of Eqs. (II.24), (II.25) is that they involve the equilibrium correlations. If we want to pursue with our graphical analysis, we thus need diagrammatic rules to represent these correlations. As we shall see soon, this will lead to further simplifications in the possible type of graphs.

To be as simple as possible, we shall limit ourselves to $\Psi_0^1(v_1;z)$; the other quantities of the theory can be handled similarly.

Putting $\rho_k^{eq} \equiv (k|\Omega^N\rho^{eq}|0)$, let us first explicitly isolate the group $(\mathscr{G})$ of particles with nonzero wave vector $k_{i\in(\mathscr{G})} \neq 0$. We write

$$\rho_k^{eq} = \tilde{\gamma}^{(\mathscr{G})}_{\{k_{i\in(\mathscr{G})}\neq 0\}} \prod_{i\notin(\mathscr{G})} \varphi^{eq}(v_i) \tag{II.26}$$

and we further decompose $\tilde{\gamma}$ in irreducible *clusters* $\gamma^{(\mathscr{G}_\alpha)}$ $(\alpha = 1\cdots n)$ according to

$$\tilde{\gamma}^{(\mathscr{G})} = \sum_{\{(\mathscr{G}_\alpha)\}} \prod_{\alpha=1}^{n} \gamma^{(\mathscr{G}_\alpha)}_{\{k_{i\in(\mathscr{G}_\alpha)}\neq 0\}}. \tag{II.27}$$

For example, we write for $\tilde{\gamma}^{(1,2,3,4)}_{k_1k_2k_3k_4}$

$$\tilde{\gamma}^{(1,2,3,4)}_{k_1k_2k_3k_4} = \gamma^{(1,2)}_{k_1k_2}\gamma^{(3,4)}_{k_3k_4} + \gamma^{(1,3)}_{k_1k_3}\gamma^{(2,4)}_{k_2k_4} + \gamma^{(1,4)}_{k_1k_4}\gamma^{(2,3)}_{k_2k_3} + \gamma^{(1,2,3,4)}_{k_1k_2k_3k_4}. \tag{II.28}$$

An important property of an irreducible cluster is that it satisfies the wave number conservation law

$$\sum_{i\in(\mathscr{G}_\alpha)} k_i = 0. \tag{II.29}$$

Each irreducible cluster will be graphically represented by a vertical bar connecting the lines $k_1, k_2, \cdots k_j$ corresponding to the clustered particles $1, 2\cdots j$. Examples are given in Fig. 3.

Thus from (II.24) we see that the graphs contributing to $\Psi_0^1(v_1;z)$ are constructed by combining in all possible ways the vertices of Fig. 1 and correlations of the type of Fig. 3 (which, if present, have of course to be put on the extreme right of the graph) with the following restrictions:

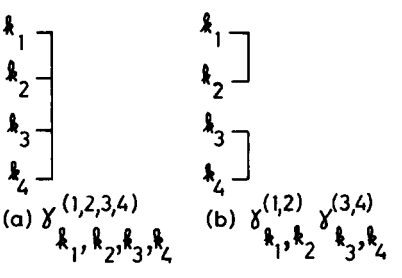


FIG. 3. Graphical representation of equilibrium correlations: (a) an irreducible cluster of fourth order (b) product of two second order irreducible clusters.

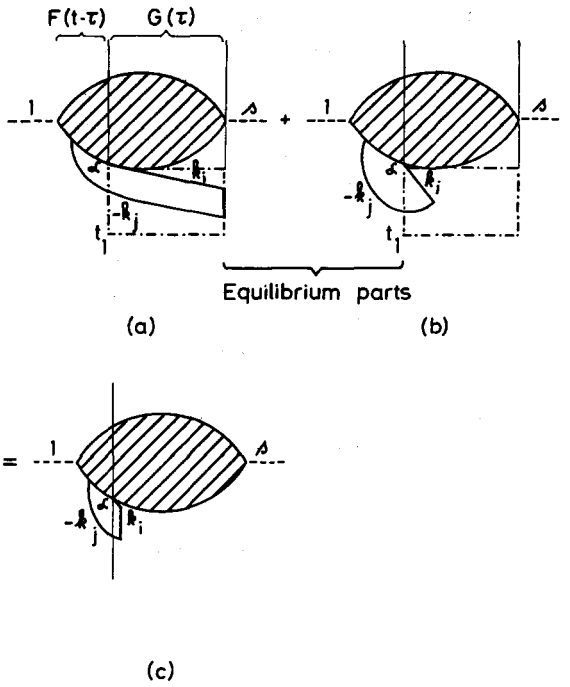
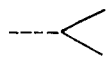


FIG. 4. Illustration of the self-propagation of the equilibrium correlations.

(1) Only fully-connected graphs are retained.

(2) The first vertex necessarily involves particle s.[16]

(3) The last vertex involves particle 1 and, as the graph ends up with $k=0$, it is of the form:

 (II. 30)

Note that, because semiconnected graphs have been eliminated, this is the *only* vertex with this topological structure which ever enters into any diagram contributing to Ψ_0^l.

We are now ready to introduce a last simplification in our diagrammatic analysis: It is based on the idea that the Liouville operator self-propagates the equilibrium correlations; more precisely, we will use the fact that

$$L\rho^{eq}=0. \tag{II. 31}$$

As a similar problem was already discussed in a different context,[17] we shall not give the general proof but rather illustrate it on a simple example; we will then formulate the general rule.

Consider the graphs (a) and (b) of Fig. 4, where the dashed structure is left arbitrary. These two graphs contain what we shall call an *"equilibrium part"* (which has been put inside a dotted rectangular frame in Fig. 4): by definition it is a part of the graph which extends up to a certain time[18] t_1 such that the particles involved in it are not connected by any line to the initial non-equilibrium particles during the time interval $0-t_1$. An equilibrium part thus describes the dynamics of equilibrium particles only and this, by (II. 31), should be trivial to treat.

That this is indeed the case in our example can be seen as follows: Using the schematic notation used in Fig. 4, we have

$$\Psi_0^{l(a)}=\int_0^t dt_1 F(t-t_1)\delta L^{\alpha i} G(t_1)[\exp i(kv_{ij}t_1)\gamma^{(i,j)}_{k_i,-k_j}] \tag{II. 32}$$

and

$$\Psi_0^{l(b)}=\int_0^t dt_1 F(t-t_1)\delta L^{\alpha i} G(t_1)\Big(\frac{1}{i}\int_0^{t_1} dt_2 \times\exp i(kv_{ij}[t_1-t_2])\delta L^{ij}\varphi^{eq}(v_i)\varphi^{eq}(v_j)\Big), \tag{II. 33}$$

where the equilibrium part only has been explicitly displayed—in both equations, it is given by the bracketed contribution.

From (II. 4) we get, after a simple calculation,

$$\Psi_0^{l(b)}=\int_0^t dt_1 F(t-t_1)\delta L^{\alpha i} G(t_1)[(1-\exp[ikv_{ij}t_1])(-\beta V_k) \times\varphi^{eq}(v_i)\varphi^{eq}(v_j)]. \tag{II. 34}$$

In (II. 33), the dynamics of the equilibrium part is treated to first order in the interaction only; for consistency, we have thus to use in (II. 32) the approximation

$$\gamma^{(i,j)}_{k_i,-k_j}\approx(-\beta V_k)\varphi^{eq}(v_i)\varphi^{eq}(v_j). \tag{II. 35}$$

We get then the remarkable, albeit not unexpected, result

$$\Psi_0^{l(a)}+\Psi_0^{l(b)}=\int_0^t dt_1 F(t-t_1)\delta L^{\alpha i} G(t_1)[(-\beta V_k)\varphi^{eq}(v_i)\varphi^{eq}(v_j)] =\int_0^t dt_1 F(t-t_1)[\delta L^{\alpha i}\gamma^{(i,j)}_{k_i,-k_j}]G(t_1). \tag{II. 36}$$

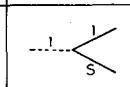
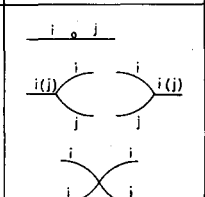
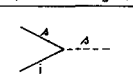
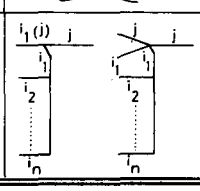
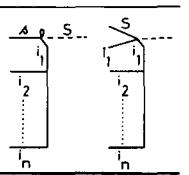
In this particular example, we see thus that equilibrium parts can be avoided (or better summed altogether) provided we use a *"renormalized vertex"*

$$\delta L^{\alpha i}\gamma^{(i,j)}_{k_i,-k_j}. \tag{II. 37}$$

This is schematically indicated in Fig. 4(c).

On the basis of (II. 31) this result can be generalized to arbitrary graphs; the diagrammatic rules for constructing Ψ_0^l become very simple: *draw all possible graphs by combining the eleven possible vertices in the way indicated in Table I.* This rule automatically insures that no semiconnected graph and no graph involving equilibrium parts will ever appear; this point is of crucial importance in the analysis of the van der Waals gas we shall present in the following sections.

TABLE I. Vertices involved in $\Psi_0^l(v_1;z)$: n is arbitrary but larger than 1.

	LAST VERTEX (extreme left)	INTERMEDIATE VERTICES	FIRST VERTICES (extreme right)
Not Renormalized			
Renormalized(*)			

(*) The contribution associated with a renormalized vertex is obtained by multiplying the contribution of the bare vertex by the equilibrium coefficient [see (II. 37)].

It is perhaps worthwhile to make here a few general remarks concerning the above discussion:

(1) Though to be complete we should first restate the rules which associate an analytical contribution to a given graph, except for the fairly simple factor associated with a renormalized vertex [compare Fig. 4(c) and Eq. (II. 37)], these rules are the same as in the graph technique of Prigogine—Balescu[9,10] and we shall thus omit doing so (see also footnote below Table I).

(2) Similarly, we should also give the corresponding analysis for the other quantities $\mathscr{C}^l_{k0}, \mathscr{D}^l_0, \mathscr{P}^l_k$ which appear in the theory. These are however easily obtained by similar arguments, and, as we shall not need these quantities explicitly in our further analysis, we shall also omit this point.

(3) We are perfectly aware that the arguments presented in this section are by no means strict proofs of our results. A general detailed proof, based on analytical consideration and not on graph analysis, is feasible (see Ref. 12 for a similar calculation) but it is unusually tedious and long. To make a complicated story as short as possible, we have preferred to present here a more intuitive argument. The interested reader can moreover check it in detail on simple examples.

Finally, to close this section, let us still perform the asymptotic integrations over τ and z in Eq. (II. 22) and (II. 23). We can follow the argument presented in Ref. 11; the result is

$$X' = -\beta n \lim_{\epsilon\to 0} \int dv_1 \Big[(0|(J^X_1 - \delta\tilde{J}^X_1)|0) + \sum_{k\neq 0} (0|(J^X_1 - \delta\tilde{J}^X_1|k) \times \mathscr{C}^l_{k0}(v_1; i\epsilon) \Big] \Big(\frac{1}{i[\Psi^l_0(v_1; i\epsilon) + i\epsilon]} \Big) \Big\{ \Big[\int dv^{N-1} \times (0|(J^X - \delta J^X)\rho^{eq}\Omega^N|0) \Big] + \mathscr{D}^l_0(v_1; i\epsilon) \Big\} \tag{II. 38}$$

and

$$X'' = -i\beta n \lim_{\epsilon\to 0} \int dv_1 \sum_{k\neq 0} (0|(J^x_1 - \delta\tilde{J}^x_1)|k)\mathscr{P}^l_k(v_1; i\epsilon). \tag{II. 39}$$

The only difference with the case of shear viscosity is the appearance of the counterterm $\delta\tilde{J}^X_1$, which exactly vanishes for this latter transport coefficient $X \equiv \eta$. However, for the example of thermal conductivity we have

$$\delta\tilde{J}^{T\kappa}_1 = \delta J^{T\kappa}_1 = (h/n)v_{1x}, \tag{II. 40}$$

while in the case of bulk viscosity, more precisely for $B = \frac{4}{3}\eta + \xi$, we have $\delta\tilde{J}^B_1 \neq \delta J^B_1 \neq 0$. As this counterterm, which plays an important role in explicit calculations, does not affect the formal discussion of the van der Waals limit, we shall postpone the proof of Eqs. (II. 38), (II. 39) until Appendix C.

III. THE γ EXPANSION OF THE LINEARIZED COLLISION OPERATOR

It follows from Eqs. (II. 38), (II. 39) that the problem of finding the γ expansion of a transport coefficient in the van der Waals limit can be reduced to the analysis of the γ dependence of the linearized quantities $\Psi^l_0(v_1; i\epsilon)$, $\mathscr{D}^l_0(v_1; i\epsilon)$, $\mathscr{C}^l_{k0}(v_1; i\epsilon)$, and $\mathscr{P}^l_k(v_1; i\epsilon)$ when the interaction potential $V(r)$ is decomposed according to (I. 1). In this section, we present such an analysis for the linearized collision operator $\Psi^l_0(v_1; i\epsilon)$.

Let us first point out that the Fourier transform of Eq. (I. 1) reads

$$V_k = V^s_k + V^L_{k\gamma\text{-}1}, \tag{III. 1}$$

where

$$V_k = \int dr V(r) \exp(ikr). \tag{III. 2}$$

If we want to apply our diagrammatic formulation to the potential (III. 1), we need to distinguish between the short- and the long-range part of this interaction: this will be accomplished by using the vertices ■ and □ for the short- and long-range contributions, respectively; lines with nonvanishing wave numbers (or dotted lines when a $k=0$ particle needs to be made explicit) will arrive at (or leave from) these vertices as they did from the "dots" used to represent the vertices in the preceding section. Moreover, this last notation will still be used whenever no distinction needs to be made between the two types of interaction.

For example, the graph of Fig. 5 will correspond to the contribution[19] [see (II. 24)]

$$\frac{n^2}{(2\pi)^3} \int dv_2 dv_3 \int d^3k \Big(V^s_k k \frac{\partial}{\partial v_{12}} \frac{1}{kv_{12} - z} V^L_{k\gamma\text{-}1} k \frac{\partial}{\partial v_{13}} \times \frac{1}{kv_{32} - z} V^s_k k \frac{\partial}{\partial v_{23}} \Big) \Phi(v_2)\varphi^{eq}(v_1)\varphi^{eq}(v_3). \tag{III. 3}$$

Notice that, owing to the presence of the long-range vertex ⊟, the integration over k in (III. 3) is restricted to the small region $|k| \leq \gamma$. This illustrates the fact that diagrams involving long-range vertices will generally contain some lines with wave vector restricted by the condition $|k| \leq \gamma$.

Now, the free long wave-length propagator

$$\frac{1}{\sum_j k_j v_j - z}, \quad |k_j| \leq \gamma \text{ (for all } j) \tag{III. 4}$$

introduces in the limit $z \to i\epsilon$ the factor γ^{-1}, as can be seen by introducing the integration variables $q_j = \gamma^{-1}k_j$ (assuming that all q_j integrals converge). As such propagation can appear an arbitrary number of times—for example, by adding more and more short range collision processes in the way illustrated in Fig. 6—it becomes clear thet *the straightforward perturbation series expansion for* $\Psi^l_0(v_1; i\epsilon)$ *contains terms which diverge in the limit* $\gamma \to 0$. (This essential point seems to be overlooked in the work of R. Elliot and L. De Sobrino).[20]

We will now show that this type of divergence can be completely eliminated if we first "renormalize" the theory by dressing the free propagators (III. 4) with all possible collision processes. Although, as we shall see later, it would be sufficient to perform this dressing to

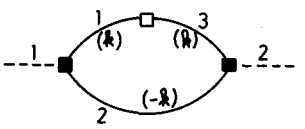


FIG. 5. An example of the decomposition of the potential in a short-and a long-range part.

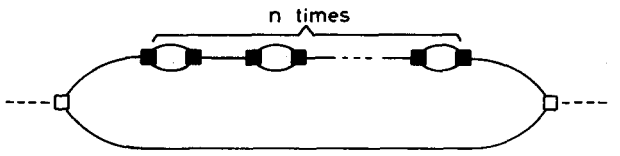


FIG. 6. Example of a class of diverging diagrams ($n=5,6\cdots$).

zeroth order in γ, it is more convenient to do it to arbitrary order in γ by defining the dressed propagator through the procedure schematically described in Fig. 7a; in this figure the dashed structure represents all possible interactions (involving *both* the short- and long-range part) starting and ending with the given wave vector k. Denoting by $\tilde{X}_k(v_1; z)$ this renormalized propagator, we have

$$\tilde{X}_k(v_1; z) = \sum_{n=0}^{\infty} \frac{1}{(kv_1 - z)} \left(\Psi_k^I(v_1; z|\gamma) \frac{1}{(kv_1 - z)} \right)^n = 1/[kv_1 - \Psi_k^I(v_1; z|\gamma) - z] \tag{III. 5}$$

Here the *inhomogeneous* linearized collision operator Ψ_k^I is the analytical contribution associated with Fig. 7b. Using the diagrammatic rules of Sec. III, it is readily verified that it can be written in complete analogy with (II. 21). We have, indeed,

$$\Psi_k^I(v_1; z|\gamma)\Phi(v_1) = \lim_{\infty} \sum_{n=1}^{\infty} \sum_{s=1}^{N} \int dv^{N-1} \left\{ \left(k^1 \right|_1 \left[-\delta L \left(\frac{1}{L_0 - z} Q_k(-\delta L) \right)^n \right]_s^{\text{F.C.}} \left| k^s \right) \times \Phi(v_s) \prod_{j \neq s} \varphi^{eq}(v_j) + \sum_{\substack{k^s \neq 0 \\ k^s = k^1}} \left(k^1 \right|_1 \left[(-\delta L) \left(\frac{1}{L_0 - z} \times Q_k(-\delta L) \right)^n \right]_s^{\text{F.C.}} \left| k^s, k' \right) (k' | \rho^{eq} \Omega^N | 0) \Phi(v_s) / \varphi^{eq}(v_s) \right\}. \tag{III. 6}$$

The notation used here is the same as in (II. 21), except for the projector Q_k which is defined by

$$Q_k = I - \sum_{a=1}^{N} |k^a)(k^a| \tag{III. 7}$$

and for the symbol $(k^i|$ [or $|k^i)$] which is used to denote a state in which particle i has wave number k, while all other particles have zero wave numbers.

Let us stress that, in Eq. (III. 6), δL involves both the short-range and the long-range part of the potential; Ψ_k^I is thus γ dependent, as is made explicit in the notation. Moreover, in the $k=0$ limit, Ψ_k^I reduces exactly to Ψ_0^I, Eq. (II. 21).

From the diagrammatic point of view, it is clear that, with the use of the renormalized propagator, we will get an exact formal expansion for the linearized collision operator Ψ_0^I by limiting ourselves to the graphs involving no "self-energy"[21] insertions.

As the physical motivation for introducing the dressed propagators (III. 5) has already been discussed in the introduction, we will not reconsider it here but we shall rather illustrate it on the very simple example shown in Fig. 8.

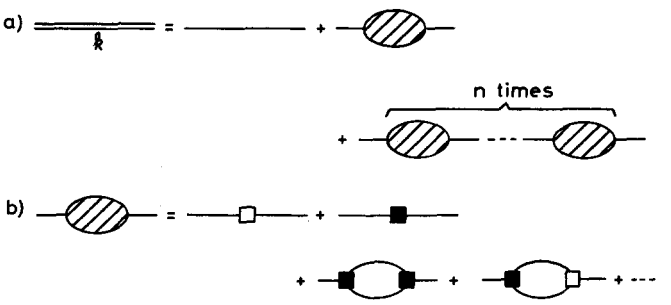


FIG. 7. The renormalized propagator: (a) general form (b) first few terms of $\Psi_k^I(v_{1t_2})$.

Using the time-dependent form of the theory,[18] and performing the change of variable $k \to q = \gamma^{-1}k$, we get from Fig. 8 the contribution

$$\lim_{\infty} \left[\frac{in\gamma^5}{(2\pi)^3} \int_0^t d\tau \int dv_2 \int d^3q \left(V_q^L q \frac{\partial}{\partial v_{12}} \right) \times \left(\frac{1}{2\pi i} \oint_c dz \exp(iz\tau) \tilde{X}_{q\gamma}(v_1; z) \right) \left(\frac{1}{2\pi i} \oint_c dz' \times \exp(-iz'\tau) \tilde{X}_{-q\gamma}(v_2; z') \right) \left(V_q^L q \frac{\partial}{\partial v_{21}} \right) [\Phi(v_1)\varphi^{eq}(v_2) + \Phi(v_2)\varphi^{eq}(v_1)] \right] \tag{III. 8}$$

to $\Psi_0^I(v_1; i\epsilon)\Phi(v_1)$.

In order to establish the γ dependence of this contribution, we have to discuss the behavior of the integral

$$X_k(v; \tau) = \frac{1}{2\pi i} \oint_c dz \exp(-iz\tau) \tilde{X}_k(v; z) \tag{III. 9}$$

in the small wave vector limit $k = \gamma q$, $\gamma \ll 1$. This problem is closely connected to the theory of linearized hydrodynamical modes developed by Résibois.[13] The detailed discussion of this important point is presented in Sec. V. We give here only the partial results which are required to continue our analysis.

The dominant contribution to (III. 9) is obtained by replacing the operator $X_k(v; \tau)$ by the sum[22]

$$\sum_{\alpha=1}^{5} \exp(\Lambda_\alpha^k(\gamma)\tau) |f_\alpha^k(\gamma)\rangle\langle \bar{f}_\alpha^k(\gamma)|. \tag{III. 10}$$

Here

$$\Lambda_\alpha^k(\gamma) = ik\Lambda_\alpha^{(1)}(k, \gamma) + k^2\Lambda_\alpha^{(2)}(k, \gamma) + O(k^3), \tag{III. 11}$$

$$\Lambda_\alpha^k(\gamma) = \Lambda_\alpha^{-k}(\gamma)^*, \tag{III. 12}$$

where $\Lambda_\alpha^{(1)}(k, \gamma)$ and $\Lambda_\alpha^{(2)}(k, \gamma)$ are real *quantities which have a finite limit when* $\gamma \to 0$. Moreover, $|f_\alpha^k(\gamma)\rangle$ and $\langle \bar{f}_\alpha^k(\gamma)|$ are, to dominant order in k, linear combinations of the invariant states of the collision operator $\Psi_0^I(v_1; i\epsilon)$; we have

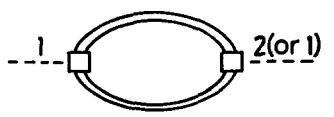


FIG. 8. A simple renormalized graph.

$$|f_\alpha^k(\gamma)\rangle = \sum_{\beta=1}^{5} c_{\alpha\beta}^{1_k}(k\gamma^{-1})\,|\beta\rangle + O(k)$$

$$\equiv |f_\alpha^{1_k}(k\gamma^{-1})\rangle + O(k), \qquad \text{(III. 13)}$$

[and a similar formula for $\langle \bar{f}_\alpha^k(\gamma)|$], where[13]

$$\langle v|1\rangle = \varphi^{eq}(v),$$
$$\langle v|i\rangle = (v_i/\sqrt{kT})\varphi^{eq}(v) \quad (i=2,3,4), \qquad \text{(III. 14)}$$
$$\langle v|5\rangle = \sqrt{2/3}\,(v^2/2kT - 3/2)\varphi^{eq}(v).$$

In Eq. (III. 12), *the coefficients* $c_{\alpha\beta}^{1_k}(k\gamma^{-1})$ have also the important property that they *tend to a finite limit when* $\gamma \to 0$; moreover, the superscript 1_k has been used in order to keep in mind that, even to zeroth order in k, the eigenmodes still depend on the direction of k.

As a matter of fact, we will prove stronger results in Sec. V, namely, we shall show that the $\Lambda_\alpha^k(\gamma)$ and $|f_\alpha^k(\gamma)\rangle$ can, respectively, be identified with the hydrodynamical eigenvalues and eigenfunctions of the van der Waals fluid; in particular, for $\gamma \to 0$, the $\Lambda_\alpha^k(\gamma)$ become simply identical to the eigenvalues (I. 2). For this reason, we shall often speak of the renormalized propagator $\tilde{X}_k(v; z)$ as describing hydrodynamical propagation in the fluid.

Substituting (III. 10) into (III. 8), we discover that the superposition of the two renormalized lines of Fig. 8 behaves, for small wave numbers $|k| = |q|\gamma$ ($\gamma \to 0$, q finite) like:

$$\sum_{\alpha,\beta=1}^{5} |\bar{f}_\alpha^{1_q}(q)\rangle|\bar{f}_\beta^{-1_q}(q)\rangle[1/i(\Lambda_\alpha^{q\gamma}(\gamma) + \Lambda_\beta^{-q\gamma}(\gamma)]\langle f_\alpha^{1_q}(q)|\langle f_\beta^{-1_q}(q)|, \qquad \text{(III. 15)}$$

so that the strongest possible divergence in the $\gamma \to 0$ limit is γ^{-2}. Consequently, the renormalized cycle of Fig. 8 represents a term of the order not less then $\gamma^5\gamma^{-2} = \gamma^3$.

We shall now extend this analysis to the general situation and determine a lower bound to the γ dependence of an arbitrary renormalized diagram, in which the free propagators have been replaced by hydrodynamical lines. For this purpose, consider any diagram which contributes to $\Psi_0^1(v_1; z)$. Taking into account the wave-number conservation law at each vertex and the limitation on these wave vectors imposed by a long-range vertex $V_{k\gamma^{-1}}^L$, namely $|k| \leq \gamma$, we divide this diagram into a sequence of neighboring segments of two types:

Type I: throughout the segment, we have hydrodynamical propagators with small wave vectors ($|k| \leq \gamma$) only.

Type II: between any two vertices of the segment, there appear some propagators carrying wave numbers not influenced by the long-range interaction, i.e., the integration over these wave vectors is γ independent and spreads over the big region $|k| \lesssim 1$ (the distances here are measured in units equal to the range of the short-range potential).

Let us stress that this division is determined by the global topological properties of the graph: If we consider a given part of a graph, we can tell whether it enters into a segment of type I or of type II only after we have determined the different wave numbers from the conservation rules in the *whole* graph.

Schematically, the division of a given diagram can be represented by drawing vertical lines through some of its vertices. A simple example is shown in Fig. 9, where the dashed structure is supposed to involve only short-range interactions.

To avoid any confusion, we adopt the following conventions: a vertex separating segment I from segment II is considered as part of the latter and the vertex appearing at the right (left) end of the diagram belongs to the corresponding right (left-) end segment.

Let us first discuss the γ dependence of type I segments. We first remark that the change of integration variable $k \to q = k\gamma^{-1}$ for all wave numbers limited to $k \leq \gamma$, introduces a γ^3 factor for each k integration.

Recalling that we always read a graph from right to left, we decide to attach each k-integration factor to the vertex where this k wave number first appears. From Table I and the definition (II. 4) of δL, we immediately get the results of Table II for the various vertices which can appear in a segment of type I. The following remarks provide useful comments about the construction of this table:

(1) According to our earlier prescriptions, we have used a dot to represent both the short- and the long-range vertices: the γ dependence is the same in both cases (see also the remark after the definition of type I and II segments).

(2) The vertex —○— is not present here, as it has been included in the definition of the renormalized propagator.

(3) The γ dependence of the renormalized vertices is a consequence of the fact that an *irreducible* cluster of n particles involves $(n-1)$ independent wave vectors.

According to the definition of a segment of type I, between any two of its neighboring vertices there appear a certain number $n \geq 2$ of hydrodynamical lines. The analysis presented for the case $n=2$ (see the example of the renormalized cycle) applies as well for $n>2$. Consequently, with any superposition of hydrodynamical lines, we shall associate the factor γ^{-2}, corresponding to the strongest possible divergence in the limit $\gamma \to 0$.

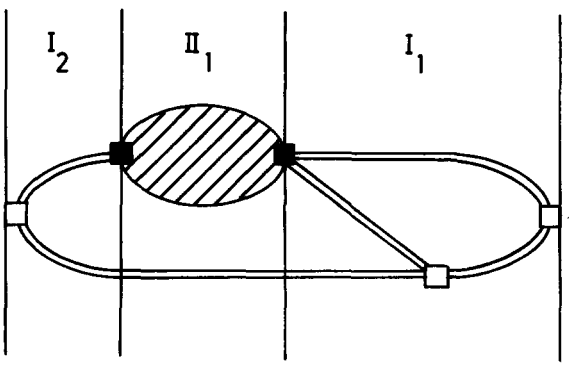


FIG. 9. Division of a diagram in segments of type I and II.

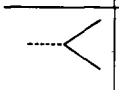
LAST VERTEX IN ψ_o^I		INTERMEDIATE VERTICES			FIRST VERTEX IN ψ_o^I	
	γ-factor		γ-factor	N_i		γ-factor
	γ		γ^4	A_i		γ^4
			γ	C_i		γ^{3n+1}
			γ^4	B_i		γ^{3n+4}
			γ^{3n+1}	$D_i^{(n)}$		
			γ^{3n+4}	$E_i^{(n)}$		

TABLE II. γ factors associated with vertices appearing in segments of type I (N_i = number of vertices in segment I_i).

After these preliminary evaluations, let us remark that the division into segments of type I and II can lead to four different structures, depending on whether the right- and left-end segments are of type I or II. Let us first consider the sequence

$$\underset{b_n\ a_n}{|I_n|}\underset{b_{n-1}\ a_{n-1}}{|\Pi_{n-1}|}\cdots\underset{b_{i+1}\ a_{i+1}}{|I_{i+1}|}\underset{b_i\ a_i}{|\Pi_i|}\cdots\underset{a_2\ b_1\ a_1}{|\Pi_1|I_1|}, \tag{III. 16}$$

where $b_i(a_i)$ equals the number of hydrodynamical lines at the left (right) extremity of segment I_i. In particular, it is convenient to define a_1 = (number of lines going *out* of the first vertex), $b_n=2$.

Applying the established rules, we can calculate the γ factor, denoted γ^{x_i}, associated with segment I_i.

Using the notation explained in Table II, we get by a straightforward calculation when $1<i<n$

$$x_i=4(A_i+B_i)+C_i+\sum_r[(3r+1)D_i^{(r)}+(3r+4)E_i^{(r)}]$$
$$-2\Big(A_i+B_i+C_i+\sum_r(D_i^{(r)}+E_i^{(r)})+1\Big). \tag{III. 17}$$

A simple calculation of the total number of lines created (or destroyed) in segment I_i leads to the identity

$$b_i-a_i=A_i-C_i+\sum_r[rD_i^{(r)}+(r+1)E_i^{(r)}], \tag{III. 18}$$

so that we get finally

$$x_i=(b_i-a_i-2)+A_i+2B_i+\sum_r[(2r-1)D_i^{(r)}+(2r+1)E_i^{(r)}]. \tag{III. 19}$$

A similar calculation can be performed when $i=1$ and $i=n$ (with $b_n=2$) and leads to

$$x_1=2(a_1-2)+b_1+A_1+2B_1+\sum_r[(2r-1)D_1^{(r)}$$
$$+(2r+1)E_1^{(r)}], \tag{III. 20}$$

$$x_n=A_n+2B_n+\sum_r[(2r-1)D_n^{(r)}+(2r+1)E_n^{(r)}]+(1-a_n). \tag{III. 21}$$

Again using (III. 18), the sum $x=\sum_{i=1}^n x_i$ can be conveniently written as

$$x=\frac{3a_1}{2}+\sum_{i=1}^{n-2}[\tfrac{3}{2}(b_i-a_i+1)-2]+\sum_{i=1}^{n}\{2B_i+(A_i+C_i)/2$$
$$+\sum_r[(\tfrac{3}{2}r-1)D_i^{(r)}+(\tfrac{3}{2}r+\tfrac{1}{2})E_i^{(r)}]\}, \tag{III. 22}$$

which gives the exponent of the factor γ^x coming from type I segments in the sequence (III. 16). One can show that this formula remains correct when $n=1$.

An analogous analysis can be performed for the three remaining sequences:

$$\underset{b_n\ a_n}{|I|}\underset{b_{n-1}}{|\Pi_{n-1}|}\cdots\underset{a_2\ b_1\ a_1}{|\Pi_1|I_1|\Pi_0|}, \tag{III. 23}$$

$$\underset{b_n\ a_n}{|\Pi_n|I_n|}\underset{b_{n-1}}{|\Pi_{n-1}|}\cdots\underset{a_2\ b_1\ a_1}{|\Pi_1|I_1|}, \tag{III. 24}$$

$$\underset{b_n\ a_n}{|\Pi_n|I_n|}\underset{b_{n-1}}{|\Pi_{n-1}|}\cdots\underset{a_2\ b_1\ a_1}{|\Pi_1|I_1|\Pi_0|}, \tag{III. 25}$$

yielding, respectively, the factors γ^{x-1}, γ^{x+b_n-3}, and γ^{x+b_n-4} where x is given by (III. 22).

Let us now estimate the influence, on the γ dependence of a diagram, coming from the segments of type II. Here, we shall be satisfied with finding the lowest order γ factor resulting from their presence. We have to distinguish between three possibilities (we take $0<i<n$):

(1) $a_{i+1}=b_i=r$: It is clear that the groups of r wave vectors $\{k_1,\ldots,k_r\}$ and $\{k_1',\ldots,k_r'\}$ which appear at the left and right side of segment Π_i must differ at least by one independent wave vector; otherwise, these lines would have propagated independently from each other, which is contrary to the definition of segment Π_i.[23] The integration over this new wave-vector yields a factor γ^3.

(2) $a_{i+1}<b_i$: In this case, we can avoid the introduction of new small wave vectors $|k|=\gamma q$ ($\gamma\to 0$, q finite) by considering segments Π_i of the type depicted in Fig. 10 (or product of disconnected structures of this type).

The left-hand extremity of segment Π_i is necessarily a vertex of the type $(k_a|\delta L^s|k_a+q_a,-q_b)$. More precisely, we can write this segment in the schematic form (the function f includes all factors which are not explicitly written!):

$$\frac{n}{8\pi^3}\int dv_b\int d^3qqV_q^s\frac{\partial}{\partial v_{ab}}\Big[(k+q)v_a-qv_b+\sum_{j\in\bar G}k_jv_j-i\epsilon\Big]^{-1}$$
$$\times f_{q_a+k_a,-q_b,\{k\}_{\bar G}}(v_a,v_b,\{v\}_{\bar G}), \tag{III. 26}$$

where $\bar G$ denotes the group of particles, different from a, which enter into region I_{i+1}. To zeroth order in γ, we can of course replace (III. 26) by

$$\frac{n}{8\pi^3}\int dv_b\int d^2qqV_q^s\frac{\partial}{\partial v_{ab}}\frac{1}{qv_a-qv_b-i\epsilon}f_{q_a,q_b,\{0\}}$$
$$\times(v_a,v_b,\{v\}_{\bar G})+0(\gamma), \tag{III. 27}$$

where due to the symmetrical role played by particles a and b, $f_{q_a,-q_b}$ is a *symmetrical* function of v_a and v_b. Equation (III. 27) is justified because all the wave vectors $\{k\}_{\bar G}$ are smaller or equal to γ by definition.

Now, according to our previous analysis, in order to get γ^{-2} factors from the hydrodynamical lines appearing in segments I_i and I_{i+1}, we have to introduce the hydro-

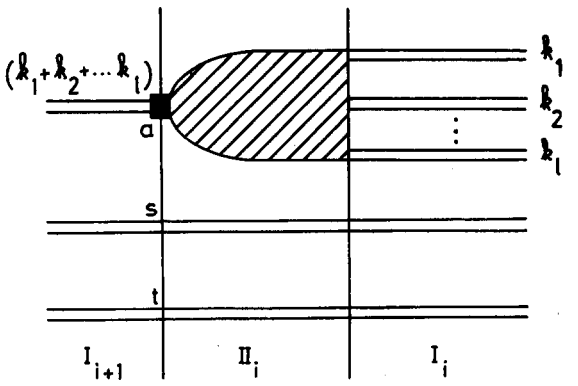


FIG. 10. Example of a Π_i segment with $a_{i+1} < b_i$.

dynamical modes $|f_\alpha^{1k}(k\gamma^{-1})\rangle$ and $\langle f_\alpha^{-1k}(k\gamma^{-1})|$ and apply Eq. (III. 10). But then segment Π_i is acted upon its left by linear combinations of the states $\langle\alpha|$ (see III. 14). Hence, to zeroth order in γ, we have to compute

$$\frac{n}{8\pi^3}\int dv_a\langle v_a|\alpha\rangle\int dv_b\int d^3q\,q\,V_q^s\frac{\partial}{\partial v_{ab}}\frac{1}{qv_{ab}-i\epsilon}f_{a,-a,\{0\}}\times(v_a,v_b,\{v\}_{\bar{G}}). \tag{III. 28}$$

Due to the symmetry of particles a and b, a simple integration by parts show that this expression vanishes. Accordingly, we conclude that the lowest order γ factor associated with this type of segment Π_i is at least γ^1.

(3) $a_{i+1} > b_i$: The group of wavevectors $\{k_1, \ldots, k_{a_{i+1}}\}$ at the left must contain at least $(a_{i+1} - b_i)$ new independent wavevectors, which do not appear in the group $\{k_1, \ldots, k_{b_i}\}$ at the right. The integration over these wavevectors leads to the factor $\gamma^{3(a_{i+1}-b_i)}$. However, we can establish a stronger result. Indeed, it is easily checked that the situation corresponding to $(a_{i+1} - b_i)$ new wavevectors can only be realized if segment Π_i "generates" the hydrodynamic propagators in the way shown in Fig. 11(a) (or by disconnected products of graphs of this same type); for example, the situation described in Fig. 11(b) introduces $(a_{i+1} - b_i + 1)$ new small wavevectors, and gives thus at least a factor $\gamma^{3(a_{i+1}-b_i+1)}$.

Now the topology depicted in Fig. 11(a) is, loosely speaking, "conjugate" to the one encountered in Fig. 10; the argument presented above can be reproduced here "mutatis mutandis", leading to an additional γ factor. Hence, in the case $a_{i+1} > b_i$, the lowest order γ factor associated with segment Π_i is $\gamma^{[3(a_{i+1}-b_i)+1]}$.

When $i = 0$ (or n), the argument presented above does not hold [because we do not have hydrodynamical modes at the right (or left)]: In this case, it is easily verified that we merely get a factor $\gamma^{3(a_1-1)}$ (or γ^0, respectively).

Finally, let us remark that the general structures (III. 16), (III. 23), (III. 24), (III. 25) which we have discussed thus far do not exhaust all the possibilities. Indeed, we should also consider the case

$$_0|\Pi_0|_0. \tag{III. 29}$$

where no hydrodynamical intermediate state can be found at all in the graph. This type of term can, however, easily be disposed of—because there is no γ^{-2}-diverging contribution coming from the propagators,

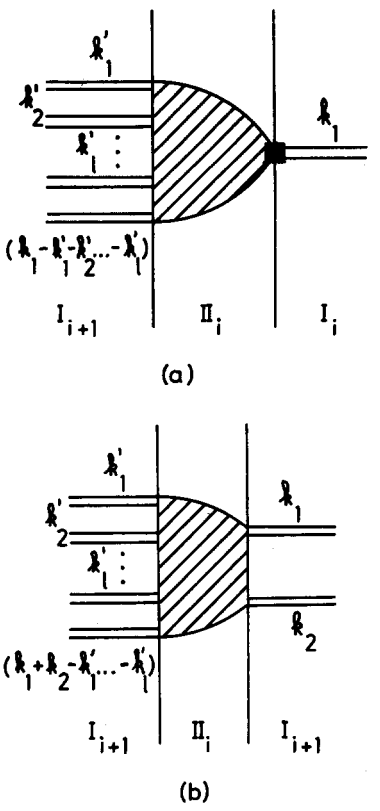


FIG. 11. Examples of Π_i segments with $a_{i+1} > b_i$.

these contributions can be expanded analytically in powers of γ. The dominant terms are, of course, of order γ^0 and correspond to the pure short-range diagrams (see example of Fig. 12a), while the addition of long range vertices introduces at least a factor γ^4 coming from (a) the restriction $|k| \leq \gamma$ over at least one wave vector integration (γ^3) and (b) a supplementary γ factor coming from the term in the expression for the long-range vertex

$$kV_{\gamma^{-1}k}^L\frac{\partial}{\partial v_{ab}} \quad (|k| \leq \gamma). \tag{III. 30}$$

For example, the graph of Fig. 12b is of order γ^4.

We can now collect these remarks to determine a lower bound for the order in γ of an arbitrary renormalized contribution to $\Psi_0^l(v_1; i\epsilon)$.

Consider first the class of diagrams whose structure corresponds to the sequence (III. 16). An elementary calculation shows us that the γ expansion of such diagram begins with terms of the order

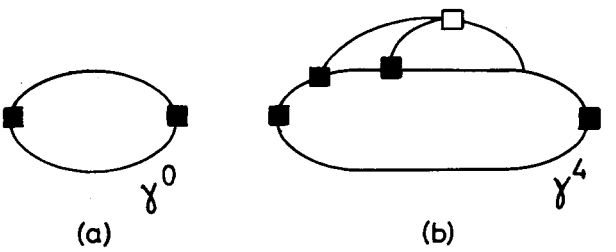


FIG. 12. Terms contributing to $_0(\Pi_0)_0$: (a) pure short range (b) graphs involving one long range interaction.

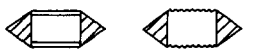
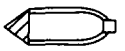
Type	Diagram	Leading order in γ
1 Eq.(III.25)		γ^1
2 Eq.(III.24)		γ^2
3 Eq.(III.23)		γ^2
4 Eq.(III.16)		γ^3

FIG. 13. Dominant corrections to the purely short-range part of Ψ_0^l in the $\gamma \to 0$ limit.

$$\Gamma = \sum_{i=1}^{n} \Big(2B_i + (A_i + C_i)/2 + \sum_r [(3r/2 - 1)D_i^{(r)} + (3r/2 + 1/2)E_i^{(r)}] \Big) + 3a_1/2 + \sum_{i=1}^{n-1} \left(\tfrac{3}{2} | b_i - a_{i+1} | + 2\delta^{k_r}_{b_i, a_{i+1}} - 1\right) . \quad \text{(III. 31)}$$

Notice that it might happen that this lowest order term vanishes when applied to a given function $\Phi(v)$; this is the reason why we call Γ a lower bound.

Similarly, we find for the minimal exponents of the remaining sequences (III. 23), (III. 24), (III. 25) the values $(\Gamma - 1)$, $(\Gamma + b_n - 3)$, and $(\Gamma + b_n - 4)$, respectively. The minimal value of the exponent Γ is also immediate to obtain; from Eq. (III. 31), it follows indeed immediately that $\Gamma_{min} = 3$, which corresponds to

$$n = 1, \quad a_1 = 2, \quad A_1 = B_1 = C_1 = D_1^{(r)} = E_1^{(r)} = 0. \quad \text{(III. 32)}$$

This result permits us to determine the structure of the lowest order diagrams in each of the four classes (III. 16), (III. 23), (III. 24), (III. 25) (in the two last cases, we have of course to put $b_n = 2$). They are represented in Fig. 13 where the dashed structure represents segment of type II, *which, to lowest order in* γ, *can be taken as purely short range* [see discussion after (III. 29)]; notice that in the first graph, which gives the dominant contribution (of order γ^1), we have to guarantee that the hydrodynamic propagators contain at least *one* long-range vertex kV^L_{kr}-1 $\partial/\partial v_{12}$, in order to insure that the corresponding k integration is restricted to $|k| \leq \gamma$ (otherwise, this term would enter into the class ${}_0|\Pi_0|_0!$). As will become clear in the course of the discussion of Sec. V, this can be achieved by substracting the terms involving purely short-range hydrodynamical lines, which we represent by a wavy line. The analytical expression associated to this purely short-range propagator is given by

$$\tilde{X}^s_k(v; z) = [kv - \Psi^{l,s}_k(v; z) - z]^{-1}, \quad \text{(III. 33)}$$

where $\Psi^{l,s}_k$ is defined as for (III. 6) with δL replaced by δL^s.

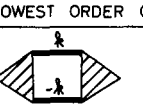
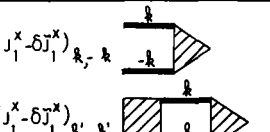
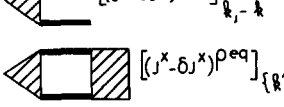
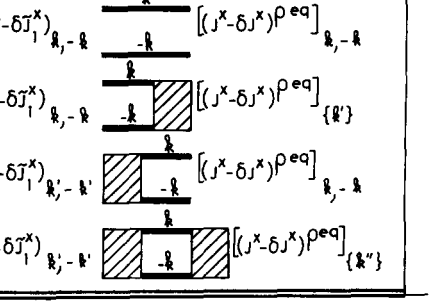
To summarize the fundamental conclusion of this analysis, we can say that the lowest order correction to the linearized homogeneous short-range collision operator due to the presence of long-range forces is of first order in γ and corresponds to diagram 1 of Fig. 13. At the same time, we proved that the hydrodynamic resummation is the proper way of analyzing the van der Waals limit $\gamma \to 0$, as it makes possible, at least in principle, the power series expansion of the collision operator in the smallness parameter γ.

Let us close this section with one more remark. In Fig. 13, we have given the dominant graphs for each of the structures (III. 16), (III. 23)—(III. 25). However, we see that the graphs 2, 3, 4 of Fig. 13 are of higher order in γ than the dominant term, given by graph 1. Except for the extremely unlikely case where this graph 1 could be shown to identically vanish up to order γ^2 (or γ^3) [for example, unexpected symmetry properties for a given function $\Phi(v_1)$, see (II. 24)], these three graphs are of a rather academic character (they can be useful in checking the general validity of our classification on simple examples) and they will not be considered any further. Indeed, in order to consistently take them into account, we should simultaneously consider the higher order corrections (i.e., γ^2 and γ^3) to graph 1. This, in particular, would imply the knowledge of the hydrodynamical eigenmodes and eigenvalues [see (III. 10)] beyond their lowest order in k and in γ: This task, although in principle possible, seems beyond the existing technical possibilities.

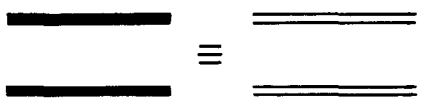
IV. EXPANSION FOR THE TRANSPORT COEFFICIENTS

Having found the lowest order diagrams contributing to $\Psi_0^l(v_1; i\epsilon)$, we can solve immediately the analogous

TABLE III. Dominant corrections to the transport coefficients (all are of order γ^1).

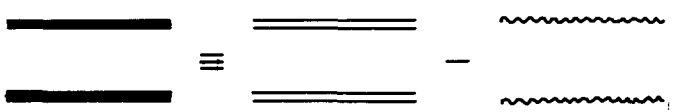
OPERATOR OR FUNCTION	LOWEST ORDER CORRECTIONS
$\Psi_0^l(i\epsilon)$	
$\sum_k (J_1^x - \delta\tilde{J}_1^x)_{k_j - k} G^l_{k\,0}(i\epsilon)$	
$\mathcal{D}_0^l(i\epsilon)$	
$\sum_k (J_1^x - \delta J_1^x)_{k_j - k} \mathcal{P}^l_k(i\epsilon)$	

problem for the operator $\mathscr{C}^l_{k0}(v_1; i\epsilon)$ and for the functions $\mathscr{D}^l_0(v_1; i\epsilon)$ and $\mathscr{P}^l_k(v_1; i\epsilon)$. Indeed, it is easy to realize that the dominant corrections to these quantities have a structure similar to those given in Fig. 13. We shall not reproduce these calculations here; the result of our classification is summarized in Table III, where the dashed structures again represent pure short-range type II segments. Moreover, we have used a shorthand notation. We have



(IV. 1a)

if (*and only if*) the long-range quantities appearing in the graph [essentially the flow operators $(J^X - \delta\tilde{J}^X)^L_{k-k}$ and $((J^X - \delta J^X)\rho^{eq})^L_{k-k}$] guarantee that the hydrodynamic propagators are limited to $|k| \leq \gamma$; while we have



(IV. 1b)

whenever there is no long-range potential—except in the hydrodynamical propagators themselves—to guarantee the condition $|k| \leq \gamma$.

To illustrate this definition, let us use the decomposition

$$(J^X - \delta\tilde{J}^X) = (J^X - \delta\tilde{J}^X)^s + (J^X - \delta\tilde{J}^X)^L \tag{IV. 2}$$

analogous to (III. 1). The graph of Fig. 14a—which is one of the dominant contribution to the creation term $(J^X - \delta\tilde{J}^X)_{k,-k}\mathscr{C}^l_{k0}(i\epsilon)$—has then to be interpreted as in Fig. 14b.

We see that in Table III, we have various matrix elements of the flow operator $(J^X - \delta\tilde{J}^X)$ and of the product $(J^X - \delta J^X)\rho^{eq}$. Let us make the following remarks about their γ-dependence:

(i) The matrix elements $(0|(J^X_1 - \delta\tilde{J}^X_1|0)$, $(0|(J^X_1 - \delta\tilde{J}^X_1)^s|k)$, and $(0|(J^X - \delta J^X)\rho^{eq}\Omega^N|0)$ are γ independent.

(ii) The only nonzero matrix elements of $(0|(J^X_1 - \delta\tilde{J}^X_1)|k)$ have the form $(0|(J^X_1 - \delta\tilde{J}^X_1)|k_1, -k_b)$. For example, in the case of thermal conductivity, we have [see (II. 7)]

$$(0|J^{T\kappa}_1|k_1, -k_b) = \frac{1}{\Omega^N}\int dr^N \exp(ikr_{1b})\left(v_{1x}\frac{1}{2}V(r_{1b}) - \frac{1}{2}v_1\frac{\partial V}{\partial r_{1b}}(r_{1b})_x\right). \tag{IV. 3}$$

For the long-range part, the term $(0|(J^x_1 - \delta\tilde{J}^x_1)^L|k_1, -k_b)$ reduces the k summation in Eqs. (II. 38), (II. 39) to the small region $|k| \leq \gamma$, while this matrix element itself is γ independent after the change of variable $k \to q\gamma$.

(iii) Similarly, we have to consider matrix elements $(k|(J^X - \delta J^X)\rho^{eq}\Omega^N|0)$. Their value corresponding to short-range interaction only will be represented by $(k|(J^X - \delta J^X)\rho^{eq}\Omega^N|0)^s$, and is of course γ independent.

Because of the long-range interaction, the matrix element $(k_a, -k_b|(J^X - \delta J^X)\rho^{eq}\Omega^N|0)$ will also contain terms confining the wavevectors k to $|k| \leq \gamma$. The lowest order term with this property (i.e., the only one which becomes γ independent after the change of variable $k \to q\gamma$) can be derived from the corresponding result for the two particle correlation function.[4] Taking again the example of thermal conductivity, we obtain

$$\begin{aligned}\int dv^{N-2}(k_a, -k_b|(J^{T\kappa} - \delta J^{T\kappa})\rho^{eq}\Omega^N|0) &= \varphi^{eq}(v_a)\varphi^{eq}(v_b)\Bigg(\Bigg\{(v_a)_x\left[\frac{v_a^2}{2} + n\left(\frac{\partial h/n}{\partial n}\right)^s_T - \frac{5kT}{2}\right] \\ &\times\Bigg\{-\frac{kT}{\Omega}\left[\left(\frac{\partial n}{\partial p}\right)^s_T\right]^2\frac{V^L_{k\gamma-1}}{1 + nV^L_{k\gamma-1}(\partial n/\partial p)^s_T}\Bigg\} + \Bigg((v_a)_x\frac{1}{2}V^L_{k\gamma-1} \\ &+ v_a\frac{1}{2}\frac{\partial kV^L_{k\gamma-1}}{\partial k_x}\Bigg)\left[kT\left(\frac{\partial n}{\partial p}\right)^s_T\frac{1}{\Omega(1 + nV^L_{k\gamma-1}(\partial n/\partial p)^s_T}\right] \\ &+ \text{idem}\, a \rightleftharpoons b\Bigg),\end{aligned} \tag{IV. 4}$$

where the thermodynamic derivatives $(\partial n/\partial p)^s_T$ and $[\partial(h/n)/\partial T]^s_n$ refer to the purely short-range system.

From Table III and Eqs. (II. 38), (II. 39), it is immediately apparent that the first correction to any transport coefficient of a van der Waals fluid is of order γ. We shall not, however, write the first correction explicitly here because it is fairly lengthy to write and because, in the present paper, we are only interested in the formal aspect of the theory. Explicit calculations will be the object of a forthcoming paper.

V. γ^0-HYDRODYNAMICAL MODES OF THE VAN DER WAALS FLUID

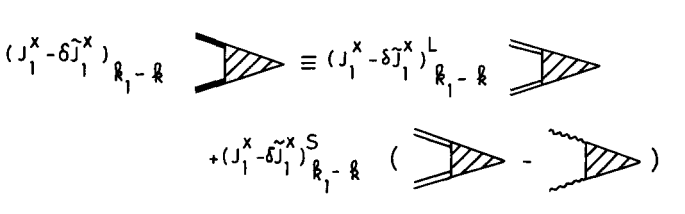
The main conclusions of the two preceding sections depend strictly upon the validity of Eqs. (III. 10)–(III. 13). We want now to justify these conjectures; moreover, in order to make future explicit calculations possible, we shall also derive the detailed form of the eigenvalues $\Lambda^k_\alpha(\gamma)$ and their corresponding eigenfunctions $|f^k_\alpha(\gamma)\rangle$ [and $\langle\bar{f}^k_\alpha(\gamma)|$] in the $\gamma \to 0$ limit.

In order to get a clue to the problem, let us first remark that if the wavelength k^{-1} was much larger than any other molecular length in the system, the considerations of Ref. 13 could be directly applied. For example, at arbitrary (*but fixed*) γ we would immediately obtain

$$\Lambda^k_{1,2}(\gamma) = \pm ikc(\gamma) - \Gamma(\gamma)k^2, \quad \Lambda^k_{3,4}(\gamma) = -\eta(\gamma)k^2/n, \quad \Lambda^k_5(\gamma) = -\kappa(\gamma)k^2/nC_p(\gamma), \tag{V. 1}$$

where

$$\Gamma(\gamma) = \frac{1}{2n}\{4\eta(\gamma)/3 + \xi(\gamma) + [1/C_v(\gamma) - 1/C_p(\gamma)]\kappa(\gamma)\}. \tag{V. 2}$$



(a) (b)

FIG. 14. Illustration of the definition (IV. 1)

Expanding thermodynamic quantities and transport coefficients in power of γ, we would then get to leading order in γ

$$\Lambda^k_{1,2}(\gamma_0) = \pm ikc(\gamma_0) - \Gamma(\gamma_0)k^2, \quad \Lambda^k_{3,4}(\gamma_0) = -\eta^s k^2/n,$$
$$\Lambda^k_5(\gamma_0) = -\kappa^s k^2/nCp(\gamma_0) \tag{V.3}$$

with

$$\Gamma(\gamma_0) = \frac{1}{2n}\{4\eta^s/3 + \xi^s + [1/C_v(\gamma_0) - 1/C_p(\gamma_0)]\kappa^s\}, \tag{V.4}$$

where $c(\gamma_0)$, $C_p(\gamma_0)$, and $C_v(\gamma_0)$ denote the well-known van der Waals limit $\gamma \to 0$ of the corresponding thermodynamic coefficients; *self-consistency* with the results of the preceding paragraph implies that the γ_0-transport coefficients simply are the pure hard-core contributions, denoted by the superscript s.

However, this result is not sufficient because in the proofs of Secs. III and IV, we have used the eigenvalues $\Lambda^k_\alpha(\gamma)$ for k smaller *or of the order* γ; k^{-1} is thus *not* much larger than the range γ^{-1} and, as a matter of fact, we have to carefully take the *double limit*:

$$\left.\begin{matrix} |k| \to 0 \\ \gamma \to 0 \end{matrix}\right\} |k|/\gamma \lesssim 1, \text{ finite.} \tag{V.5}$$

As we shall see now, the analysis in this limit is very delicate although it leads to remarkably simple results. Let us consider the operator

$$\bar{\Psi}^l_k(v; z\,|\gamma) = [-kv_x + \Psi^l_k(v; z\,|\gamma)], \tag{V.6}$$

where, to simplify, we have oriented the wave number k along the x axis.

We want to determine those eigenvalues $\Lambda^k_\alpha(\gamma)$ and eigenfunctions $|f^k_\alpha(\gamma)\rangle$ of the problem

$$[\bar{\Psi}^l_k(v_1; i\Lambda^k_\alpha(\gamma)\,|\gamma) + i\Lambda^k_\alpha(\gamma)]|f^k_\alpha(\gamma)\rangle = 0, \tag{V.7}$$

such that $\Lambda^k_\alpha(\gamma) \to 0$ in the double limit (V.5). Note that it is not immediately obvious that such eigenvalues do exist (because the ratio $k\gamma^{-1}$ is kept fixed). However, our present analysis will show that there are indeed five such modes.

In order to make the limiting procedure (V.5) explicit, we shall write our equations in terms of the variables

$$\gamma \text{ and } y = |k|\gamma^{-1}, \tag{V.8}$$

where y is kept fixed while $\gamma \to 0$.

Setting (for simplicity, we drop the velocity dependence on the lhs)

$$\Psi_\gamma(y; z) \equiv \bar{\Psi}^l_{\gamma y}(v; z\,|\gamma),$$

we have thus to look for the expansion

$$\Psi_\gamma(y; z) = \Psi^{(0)}(y, z) + \gamma\Psi^{(1)}(y, z) + \gamma^2/2\Psi^{(2)}(y, z) + O(\gamma^3) \tag{V.9}$$

from which the eigenvalues (II.7), now written $\bar{\Lambda}^\gamma_\alpha(y)$, will be easy to calculate by perturbation method to second order in γ.

In order to carefully analyze the behavior of $\Psi_\gamma(y, z)$, we shall again use our diagram technique.

Using the similarity between Ψ^l_k and Ψ^l_0 [to which the former reduces when $k = 0$; see (II.27) and (III.6)], it is immediate to construct the vertices from which Ψ^l_k, and thus $\Psi_\gamma(y; z)$, is built. One readily verifies that the intermediate vertices are the same as in Table I, while the first and last vertices are different (the ingoing and outgoing line now carry the nonvanishing wave vector k) and are given in Table IV. Note the very important point that a *new* first vertex appear (the starred vertex of Table IV) in which the ingoing particle s completely exchanges its wave vector with a correlated particle i. This vertex vanishes identically when $k = 0$. Moreover, we also have another graph which does not appear in the homogeneous limit, namely the first order Vlasov term, shown in Fig. 17(a).

In order to systematically construct the expansion (V.10), it is convenient to separately consider four categories of terms:

(1) The first, and most trivial one, is simply the pure hard-core term which we denote by the superscript [1] in brackets,

$$[\Psi_\gamma(y, z)]^{[1]} \equiv \bar{\Psi}^{l,s}_{\gamma y}(v; z), \tag{V.10}$$

which of course only depends on $k = \gamma y$ and leads thus to the expansion

$$[\Psi_\gamma(y, z)]^{[1]} = \tilde{\Psi}^{(0)}(z) + \gamma y\tilde{\Psi}^{(1)}(z) + [(\gamma y)^2/2!]\tilde{\Psi}^{(2)}(z), \tag{V.11}$$

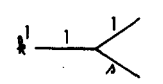
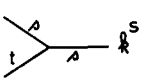
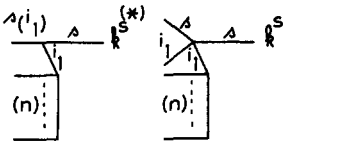
where $\tilde{\Psi}^{(i)}$ is y independent and is defined by

$$\tilde{\Psi}^{(i)}(z) = \left(\frac{\partial^i \bar{\Psi}^{l,s}_{k'}}{\partial k^i}\right)_{k=0}. \tag{V.12}$$

(2) The next group of terms by definition involves at least one long range interaction and is built with all possible vertices *except* the starred vertex of Table IV; the only exception is the pure Vlasov term (see Fig. 7) which we put in category 3 below. As example is given in Fig. 15 and its dominant contribution is, after propagator renormalization[24]:

$$\frac{in}{(8\pi^3)}\int dv_2 \int d^3k' k' V^s_{k'} \frac{\partial}{\partial v_1}\left(\sum_{\alpha,\alpha'}[1/(\Lambda^{k'}_\alpha + \Lambda^{-k'+k}_{\alpha'})]\right.$$
$$\times |f^1_\alpha{}^{k'}(k'\gamma^{-1})\rangle |f^1_\alpha{}^{(k-k')}(|k-k'|\gamma^{-1})\rangle\langle\bar{f}^1_\alpha{}^{k'}(k'\gamma^{-1})$$
$$\left.\langle\bar{f}^1_\alpha{}^{(k-k')}(|k-k'|\gamma^{-1})|\right)(k+k')V^L_{\gamma^{-1}(k+k')}\frac{\partial}{\partial v_{12}}\Phi(v_2)\varphi^{eq}(v_1). \tag{V.13}$$

TABLE IV. First and last vertex of Ψ^l_k.

LAST VERTEX (EXTREME LEFT)	FIRST VERTEX (EXTREME RIGHT)
	
	

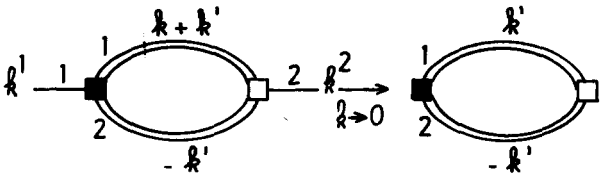


FIG. 15. A contribution to Ψ_k^l which reduces to Ψ_0^l when $k \to 0$.

Changing to the variable $k' \to q'\gamma$, we get

$$\frac{in\gamma^3}{8\pi^3}\int dv_2 \int d^3q' q' V^s_{q'\gamma}\frac{\partial}{\partial v_1}\left(\sum_{\alpha\alpha'}\gamma^2/(\Lambda^{q'\gamma}_\alpha + \Lambda^{|q'+y|\gamma}_{\alpha'})f^1_\alpha{}^{q'}(q)\rangle \times |f^1_\alpha{}^{(q'-y)}(|q'-y|)\rangle\langle\bar{f}^1_\alpha{}^{q'}(q)|\langle\bar{f}^1_\alpha{}^{(q'-y)}(|q'-y|)|\right) \times (q'+y)V^L_{q'+y}\frac{\partial}{\partial v_{12}}\Phi(v_2)\varphi^{eq}(v_1) \propto \gamma^3. \quad (V.14)$$

Let us then remark that if we *formally* let $y \to 0$ in Eq. (V.14), this contribution goes smoothly into a contribution of the homogeneous operator Ψ_0^l (see Fig. 15b; moreover, *the γ dependence is unchanged by this limiting procedure.*

This remark is general and, by this trick, we can readily construct the graphs of category (2) which are of order γ and γ^2. (None of them is of order γ^0 as a consequence of our analysis of Sec. III.) They are indicated in Fig. 16. Denoting them, respectively, by $\gamma\delta_1\Psi_\gamma$ and $\gamma^2\delta_2\Psi_\gamma$, we have thus

$$[\Psi_\gamma(y;z)]^{[2]} = \gamma\delta_1\Psi_\gamma(y;z) + \gamma^2\delta_2\Psi_\gamma(y;z) + O(\gamma^3). \quad (V.15)$$

This result can be further expanded to second order in γ and thus leads to

$$[\Psi_\gamma(y;z)]^{[2]} = \gamma\delta_1\Psi_0(y;z) + \frac{\gamma^2}{2}\left(2\frac{\partial\delta_1\Psi_\gamma(y;z)}{\partial\gamma}\Big|_{\gamma=0} + 2\delta_2\Psi_0(y;z)\right) + O(\gamma^3). \quad (V.16)$$

(3) The third category of graph is most important; it is made of the Vlasov term and of all contributions which involve the *long-range* starred vertex of Table IV. These graphs are illustrated in Fig. 17.

It is convenient to put these terms together because they have a similar structure. Indeed, the Vlasov term reads

$$-\eta\beta k v_1 V^L_{k\gamma-1}\varphi^{eq}(v_1)\int dv_2\Phi(v_2) \equiv -\gamma n\beta y v_1 V^L_y\varphi^{eq}(v_1)\int dv_2\Phi(v_2), \quad (V.17)$$

while the graphs involving the starred vertex of Table II can be written generally as (see III. 6)

$$\sum_{n=1}^{\infty}\sum_{i=1}^{N}\int dv^{N-1}\sum_{\substack{k_i'\neq 0,\\ k_s'=0}}\left(k_1\left|\left[\left(-\delta L Q_k\frac{1}{L_0-z}\right)^n\right]^{F.C.}_i\right|k_i+k_i',k'\right) \times[-n\beta k v_i V^L_{k\gamma-1}\int d^3v_s\Phi(v_s)](k'|\rho^{eq}\Omega^N|0)/\varphi^{eq}(v_s). \quad (V.18)$$

Using the abbreviation

$$\mathscr{D}_{k^1;k',k_i} \equiv \sum_{n=1}^{\infty}\left(k^1\left|\left[\left(-\delta L Q_k\frac{1}{L_0-z}\right)^n\right]^{F.C.}_i\right|k_i+k_i',k'\right),$$

we can combine Eq. (V.17) and (V.18) to get

$$[\Psi_\gamma(y;z)]^{[3]}\Phi(v_1) = \gamma\sum_{\substack{i=1\\ i\neq s}}^{N}\left[[-n\beta y V^L_y\int dv_s\Phi(v_s)] \times\left(v_1\varphi^{eq}(v_1)\delta^{K_\gamma}_{1,i} + \int dv^{N-1} \times\sum_{\substack{k_i'\neq 0\\ k_s'=0}}\mathscr{D}_{\gamma y^1;k',\gamma y_i}v_i\rho^{eq}_{k'}/\varphi^{eq}(v_s)\right)\right]. \quad (V.19)$$

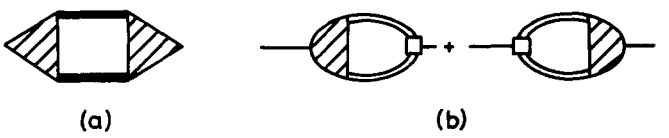


FIG. 16. Contributions to Ψ_k^l of category 2: (a) dominant order γ: $\delta_1\Psi_\gamma(y;z)$ (b) dominant order γ^2: $\delta_2\Psi_\gamma(y;z)$.

Now the so-called destruction operator $\mathscr{D}$, as well as the equilibrium correlation $\rho^{eq}_{k'}$, still involve an arbitrary number of long-range interactions. They have thus to be classified in powers of γ as we have done for the graphs of category 3; this again can be achieved with the method of Sec. III—the leading term, of order γ^0 [except for the $k=\gamma y$ dependence explicitly written in (V.19)] is of course the pure hard-core contribution. We have, moreover, a γ contribution, which, combined with the γ factor of Eq. (V.19), leads to a γ^2 term and should thus be retained in the expansion (V.11). We have, however, shown that this term does not contribute to the eigenvalues $\bar{\Lambda}^\gamma_\alpha(y)$ at order γ^2. As a similar situation will be explicitly dealt with later, in conjunction with the terms $\delta_1\Psi_\gamma(y;z)$ and $\delta_2\Psi_\gamma(y;z)$ of Eq. (II. 16) (see Appendix A), we shall not reproduce this calculation here. We can thus replace Eq. (V.19) by a similar equation with the change

$$\mathscr{D}_{\gamma y^1;k',\gamma y_i} \to \mathscr{D}^s_{\gamma y^1;k',\gamma y_i}, \quad \rho^{eq}_{k'} \to \rho^{eq\,s}_{k'}. \quad (V.20)$$

Notice that, although we now only have pure hard-core terms in (V.20), we can still expand $\mathscr{D}^s$ in power of γ according to

$$\mathscr{D}^s_{\gamma y^1;k',\gamma y_i} = \mathscr{D}^s_{0;k',0} + \gamma y\left(\frac{\partial}{\partial k}\mathscr{D}^s_{k';k',k_i}\right)_{k=0}. \quad (V.21)$$

(4) The last category of terms involves the short-range starred vertex of Table IV, together with at least one long-range vertex of another type. Here again, one can show, following the method of Sec. III, that there is indeed a γ^2 contribution of this type but, again, it does not contribute to the eigenvalues $\bar{\Lambda}^\gamma_\alpha(y)$. We shall then simply put

$$[\Psi_\gamma(y;z)]^{[4]} = O(\gamma^3). \quad (V.22)$$

Having thus exhausted all possible contributions to $\Psi_\gamma(y;z)$, we see immediately, by inserting (V.11), (V.16), (V.19) and (V.22) into (V.10), that the only γ independent part of this operator is the pure hard-core

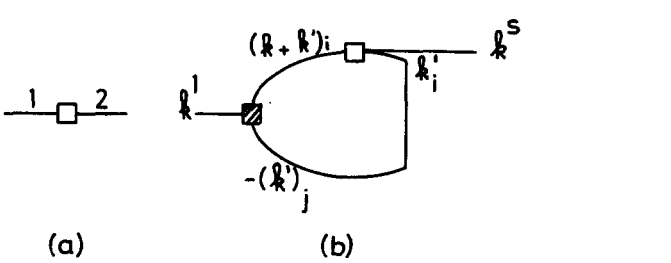


FIG. 17. Examples of contributions to $\Psi_\gamma(y;z)$ of category 3.

term $\Psi_0^{l,s}(v_1; z)$. As, in the limit $z \to 0$, we know that this operator conserves the five eigenvectors (III. 14), we are thus by now sure that there indeed exists five eigenvalues which tend to zero when $\gamma \to 0$; these are the five hydrodynamical modes of the van der Waals fluid.

We should now explicitly construct these modes, following closely the method of Ref. 13. As this calculation involves no really new feature, we leave for Appendix A a brief account of this proof and we here merely present the results.

In the limit $\gamma \to 0$, y finite (or $y \to 0$), the five hydrodynamical eigenvalues $\Lambda_\alpha^k(\gamma)$ are given by Eqs. (I. 2), (I. 3), up to terms of order γ^3. In these equations, we have

$$c(k\gamma^{-1}) = \left(\frac{C_p(k\gamma^{-1})}{C_v(k\gamma^{-1})} \frac{n^{-1}}{\chi_T(k\gamma^{-1})}\right)^{1/2}, \tag{V.23}$$

$$C_p(k\gamma^{-1}) = C_p^s + \frac{T}{n}\left[\left(\frac{\partial p}{\partial T}\right)_n^s\right]^2 [\chi_T(k\gamma^{-1}) - \chi_T^s], \tag{V.24}$$

$$C_v(k\gamma^{-1}) \equiv C_v^s, \tag{V.25}$$

where we have put

$$\chi_T(k\gamma^{-1}) = \frac{n^{-1}}{(\partial p/\partial n)_T^s + nV_y^L}. \tag{V.26}$$

These quantities respectively represent the finite wave number generalization of the sound velocity, specific heat at constant pressure and volume, and compressibility of a van der Waals fluid. In the limit $y = k\gamma^{-1} \to 0$, they reduce to the well-known corresponding thermodynamic quantities.[4,5]

The eigenmodes $|f_\alpha^{1k}(k\gamma^{-1})\rangle$ and $\langle \bar{f}_\alpha^{1k}(k\gamma^{-1})|$ associated to these eigenvalues are given by the following formulas, valid to order γ^0 (1_k is assumed to be along the x axis):

$$|f_{1,2}^{1k}(k\gamma^{-1})\rangle = \frac{1}{\sqrt{2}}\left[\frac{\sqrt{kT}}{c(k\gamma^{-1})}|1\rangle \pm |2\rangle + \frac{1}{nc(k\gamma^{-1})C_v^s} \times \sqrt{3kT/2}\left(\frac{\partial p}{\partial T}\right)_n^s |5\rangle\right], \tag{V.27a}$$

$$|f_{3,4}^{1k}(k\gamma^{-1})\rangle = |3,4\rangle, \tag{V.27b}$$

$$|f_5^{1k}(k\gamma^{-1})\rangle = \frac{1}{c^2(k\gamma^{-1})}\left[-\sqrt{\tfrac{2}{3}}\frac{T}{n}\left(\frac{\partial p}{\partial T}\right)_n^s |1\rangle + \frac{n^{-1}}{\chi_T(k\gamma^{-1})}|5\rangle\right], \tag{V.27c}$$

while

$$\langle \bar{f}_{1,2}^{1k}(k\gamma^{-1})| = \frac{1}{\sqrt{2}}\left[\frac{1}{c(k\gamma^{-1})\sqrt{kT}}\frac{n^{-1}}{\chi_T(k\gamma^{-1})}\langle 1| \pm \langle 2| + \sqrt{2kT/3}\,\frac{1}{c(k\gamma^{-1})nk}\left(\frac{\partial p}{\partial T}\right)_n^s \langle \bar{\chi}_5|\right], \tag{V.28a}$$

$$\langle \bar{f}_{3,4}^{1k}(k\gamma^{-1})| = \langle 3,4|, \tag{V.28b}$$

$$\langle \bar{f}_5^{1k}(k\gamma^{-1})| = \left[-\frac{1}{nC_v^s}\sqrt{3/2}\left(\frac{\partial p}{\partial T}\right)_n^s \langle 1| + \langle \bar{\chi}_5|\right], \tag{V.28c}$$

where we have introduced the bra $\langle \bar{\chi}_5|$ defined by

$$\langle \bar{\chi}_5| = \frac{2k}{3C_v^s}\left\{\langle 5| - (\tfrac{2}{3})^{1/2}\frac{1}{kT}\left[\left(\frac{\partial e}{\partial n}\right)_T^s - \frac{3kT}{2} - nV_0^s\right]\langle 1|\right\}. \tag{V.29}$$

In these equations, we have taken (V. 25) into account, as well as the analogous results

$$\left(\frac{\partial p(k\gamma^{-1})}{\partial T}\right)_n = \left(\frac{\partial p}{\partial T}\right)_n^s, \tag{V.30}$$

$$\left(\frac{\partial e(k\gamma^{-1})}{\partial n}\right)_n = \left(\frac{\partial e}{\partial n}\right)_T^s + nV^L(k\gamma^{-1}) \tag{V.31}$$

which are valid in the $\gamma \to 0$, $k\gamma^{-1}$ finite limit.

Let us also point out an interesting orthogonality property of the $|f_\alpha^{1k}\rangle$ and $\langle \bar{f}_\alpha^{1k}|$:

$$\langle \bar{f}_\alpha^{1k}(k\gamma^{-1})|\left(1 + \frac{\partial \Psi_0^{l,s}}{\partial z}\bigg|_{z=0}\right)|f_{\alpha'}^{1k}(k\gamma^{-1})\rangle = \delta_{\alpha,\alpha'}^{Kr}. \tag{V.32}$$

This property is proved in Appendix A, together with a sketch of the proof of Eqs. (V. 23) to (V. 29). Moreover, in Appendix B, we show that Eq. (III. 10) is a direct consequence of our definition of the eigenfunctions $|f_\alpha^{1k}\rangle$, $\langle \bar{f}_\alpha^{1k}|$ and eigenvalues $\Lambda_\alpha^k(\gamma)$.

VI. CONCLUDING REMARKS

In Sec. III, we have studied the effect of the long-range forces on the linearized collision operator $\Psi_0^l(v_1; i\epsilon)$. We have seen that the correct treatment of the problem requires the concept of microscopic hydrodynamical modes of the van der Waals fluid for wave number $k \leq \gamma$. Assuming the existence of such modes, we were able to find the lowest order corrections to this collision operator. In Sec. IV the proper generalization to the calculation of an arbitrary transport coefficient of a pure fluid[25] was made; the formal result is given in Table III. Finally, in Sec. V, we have justified the existence of such modes for $k \leq \gamma$; by a fairly complicated argument, we arrived at a remarkably simple result: These modes are defined, to leading order in γ, in terms of the purely short-range transport coefficients and the suitably generalized thermodynamic coefficients of the van der Waals fluid.

Let us still make two general remarks about these modes:

(1) Formulas (I. 2) (I. 3) and (V. 23), (V. 25) give expressions for the eigenvalues $\Lambda_\alpha^k(\gamma)$ identical to those implicit in the work of Kawasaki[8] and others in their discussion of the critical transport properties of a van der Waals fluid. There, these eigenvalues result from the macroscopic hydrodynamical equations of a hard-core fluid, suitably modified by an average field term, describing the effect of the long-range forces. Sec. V thus gives the proper microscopic justification to Kawasaki procedure, far from the critical point at least.

(2) The calculations of Sec. V is based on self-consistency: We assume the existence of the modes in order to construct them explicitly. In some sense, we have only proved here the sufficiency of this self-consistency. Although this would be very tedious, there should be no difficulty to prove the necessary character of this self-consistency; we would merely have to assume that the modes give a dominant contribution of order γ^{-n} (where n is left unspecified) and show then that we indeed have $n = 2$. Moreover, self-consistency of our method is also supported by the result of Sec. IV

(where indeed the γ^0-transport coefficients were shown to be of a purely short-range nature) and by Eqs. (V. 1), (V. 2), to which our eigenvalues $\Lambda_\alpha^k(\gamma)$ reduce when $k\gamma^{-1} \to 0$.

(3) The main results of this paper, which are summarized in Table III. indicate that the lowest order correction to any transport coefficient has the structure of a mode—mode coupling term, in which two long wavelength propagators interact through short-range processes and mechanisms involving the flows ($J^K - \delta J^K$). These results are valid to an arbitrary order in the density and, within the frame of formal perturbation theory, they are exact. Yet, in the form presented in Table III, they are still quite far from explicit. Due to special properties of the short-range collision operator in the long wavelength limit, it is however possible to considerably simplify these expressions; they lead then to simple quadratures, involving only the long-range potential $V^L_{k\gamma^{-1}}$ and the macroscopic properties of the short-range reference fluid. This remarkable result will be the object of a forthcoming publication.

ACKNOWLEDGMENTS

The first thoughts to this problem were given when one of the authors (P.R.) was visiting Professor J. Lebowitz at Yeshiva University, with the financial help of the AFOSR Grant No. 68-1416-A. One of us (J.P.) wishes to acknowledge his gratitude to Professor Prigogine for his kind hospitality at the University of Brussels.

APPENDIX A[26]

We want to sketch here how the hydrodynamical modes of the van der Waals fluid, introduced in the main text, can be obtained from the γ expansion (V. 9) of the eigenvalue problem (V. 7). The method used here will be slightly different from that of Refs. 13, 27, and 28, (hereafter referred to as I, II, III); however, in intermediate steps we shall often use results derived in these articles.

Because we look only for hydrodynamical eigenvalues $\bar{\Lambda}^\gamma_\alpha(y)$ up to order γ^2, we can transform Eq. (V. 7) in [see also (V. 8), (V. 9)]

$$i[\Psi_\gamma(y) - \tfrac{1}{2}\ddot{\Psi}_\gamma(y)(\bar{\Lambda}^\gamma_\alpha(y))^2]|f^\gamma_\alpha(y)\rangle = [1 + \dot{\Psi}_\gamma(y)]\bar{\Lambda}^\gamma_\alpha(y)|f^\gamma_\alpha(y)\rangle, \tag{A. 1}$$

where the dot indicates a derivative with respect to z taken at $z = +i\epsilon \to 0$; for example,

$$\Psi_\gamma(y) = \Psi_\gamma(y; z=0), \quad \dot{\Psi}_\gamma(y) = \frac{\partial \Psi_\gamma(y; z)}{\partial z}\bigg|_{z=0}, \text{ etc.} \tag{A. 2}$$

Using (V. 10) and γ expansions for the eigenvalues $\bar{\Lambda}^\gamma_\alpha(y)$ $[\alpha \in (G); (G) = 1, 2 \cdots 5]$ and eigenfunctions $|f^\gamma_\alpha(y)\rangle$, we have

$$\bar{\Lambda}^\gamma_\alpha(y) = \gamma\bar{\Lambda}^{(1)}_\alpha(y) + \gamma^2\bar{\Lambda}^{(2)}_\alpha(y) + \cdots, \tag{A. 3}$$

$$|f^\gamma_\alpha(y)\rangle = |f^{(0)}_\alpha(y)\rangle + \gamma|f^{(1)}_\alpha(y)\rangle + \gamma^2|f^{(2)}_\alpha(y)\rangle + \cdots. \tag{A. 4}$$

Identifying equal powers of γ, we get from (A. 1)

$$i\Psi^{(0)}|f^{(0)}_\alpha(y)\rangle = 0, \tag{A. 5a}$$

$$i(\Psi^{(1)}(y)|f^{(0)}_\alpha(y)\rangle + \Psi^{(0)}|f^{(1)}_\alpha(y)\rangle) = [1 + \dot{\Psi}^{(0)}]\bar{\Lambda}^{(1)}_\alpha(y)|f^{(0)}_\alpha(y)\rangle, \tag{A. 5b}$$

$$i(\tfrac{1}{2}\Psi^{(2)}(y)|f^0_\alpha(y)\rangle + \Psi^{(1)}(y)|f^{(1)}_\alpha(y)\rangle + \Psi^{(0)}|f^{(2)}_\alpha(y)\rangle) - i\tfrac{1}{2}\ddot{\Psi}^{(0)}(y)(\bar{\Lambda}^{(1)}_\alpha(y))^2|f^{(0)}_\alpha(y)\rangle - (1 + \dot{\Psi}^{(0)})\bar{\Lambda}^{(1)}_\alpha(y)|f^{(1)}_\alpha(y)\rangle - \dot{\Psi}^{(1)}(y)\bar{\Lambda}^{(1)}_\alpha(y)|f^{(0)}_\alpha(y)\rangle = (1 + \dot{\Psi}^{(0)})\bar{\Lambda}^{(2)}_\alpha(y)|f^{(0)}_\alpha(y)\rangle. \tag{A. 5c}$$

In these equations, we have taken into account that $\Psi^{(0)}$ is made of purely short-range contributions; it is thus y independent.

Let us study these equations in succession:

(1) *Order* γ^0: One verifies immediately [see I, Eq. (79)] that the five eigenfunctions of (A. 5a) are precisely given by $|\alpha\rangle$ $[\alpha \in (G)]$, defined in Eq. (III. 14). However, in order to set up a perturbation scheme, we need a complete basis and it is natural to take this basis from the more general eigenvalue problem[29]

$$i\Psi^{(0)}|\chi_n\rangle = (1 + \dot{\Psi}^{(0)})\mu^0_n|\chi_n\rangle \tag{A. 6}$$

and from the adjoint problem

$$\langle\bar{\chi}_n|i\Psi^{(0)} = \langle\bar{\chi}_n|(1 + \dot{\Psi}^{(0)})\mu^0_n. \tag{A. 7}$$

Taking into account the fact that $i\Psi^{(0)}$ and $\dot{\Psi}^{(0)}$ are Hermitian operators, we see that the eigenvalues are all real. Moreover, for distinct eigenvalues $\mu^0_n \neq \mu^0_m$, the standard procedure shows that the corresponding eigenfunctions are orthogonal with the weight $(1 + \dot{\Psi}^{(0)})$:

$$\langle\bar{\chi}_n|(1 + \dot{\Psi}^{(0)})|\chi_m\rangle = 0 \quad (\mu^0_n \neq \mu^0_m), \tag{A. 8}$$

where the scalar product $\langle f|g\rangle$ is defined by

$$\langle f|g\rangle = \int dv\varphi^{eq}(v)^{-1}f^*(v)g(v). \tag{A. 9}$$

For nondegenerate eigenvalues, hermiticity also insures that $\langle\bar{\chi}_n| \equiv (|\chi_n\rangle)^\dagger$ and one can thus normalize these eigenfunctions according to

$$\langle\chi_n|(1 + \dot{\Psi}^{(0)})|\chi_n\rangle = 1. \tag{A. 10}$$

For degenerate eigenvalues, one can still orthonormalize the eigenfunctions according to (A. 8) and (A. 9) by the Schmidt orthogonalization procedure and we shall assume that this has been done for the eigenvalues $n \notin (G)$. Yet, in the case $n \in (G)$, contact with the previous work of I, II, III will be easier if we choose different right and left eigenfunctions. We take as right eigenfunctions the $|\alpha\rangle$, given by Eq. (III. 4):

$$|\chi_\alpha\rangle \equiv |\alpha\rangle. \tag{A. 11}$$

These functions are of course orthonormal with respect to the norm defined with the scalar product (A. 9), but not with respect to (A. 8), (A. 10). The left eigenfunctions $\langle\bar{\chi}_\alpha|$ have then to be constructed from the linear combination of the $\langle\alpha|$ such that

$$\langle\bar{\chi}_\alpha|(1 + \dot{\Psi}^{(0)})|\chi_\beta\rangle = \delta^{Kr}_{\alpha,\beta} \quad [\alpha, \beta \in (G)]. \tag{A. 12}$$

From the matrix elements calculated in I [see I, Eq. (87)],[30] one immediately obtains

$$\langle\bar{\chi}_\alpha| = \langle\alpha| \quad (\alpha = 1, 2, 3, 4), \tag{A. 13a}$$

$$\langle\bar{\chi}_5| = \frac{3k}{2C_v}\left\{\langle 5| - \sqrt{2/3}\,\frac{1}{kT}\left[\left(\frac{\partial e}{\partial n}\right)^s_T - \frac{3kT}{2} - nV_0^s\right]\langle 1|\right\}. \tag{A. 13b}$$

We furthermore assume that the set $\langle\bar{\chi}_n|$, $|\chi_n\rangle$ (for *all* n) which satisfies the orthogonality condition

$$\langle\bar{\chi}_n|(1+\dot{\Psi}^{(0)})|\chi_m\rangle=\delta^{Kr}_{n,m} \quad \text{(for all } n\text{)} \tag{A.14}$$

forms a complete set. The completeness relation takes then the unusual form

$$\sum_n (1+\dot{\Psi}^{(0)})|\chi_n\rangle\langle\bar{\chi}_n| = \sum_n |\chi_n\rangle\langle\bar{\chi}_n|(1+\dot{\Psi}^{(0)})=1. \tag{A.15}$$

Finally, we will also assume that zero is not an accumulation point of the $\{\mu_n^0\}$.

(2) *Order* γ: The degeneracy of the unperturbed eigenvalues associated with the $|\chi_\alpha\rangle$ ($\langle\bar{\chi}_\alpha|$) prevents us from immediately performing a perturbation calculus of the eigenvalues $\bar{\Lambda}^\gamma_\alpha$. We first have to remove this degeneracy by solving exactly the first order problem within the subspace spanned by these eigenfunctions. Let us put

$$|f_\alpha^{(0)}\rangle=\sum_\beta c_{\alpha\beta}|\chi_\beta\rangle, \tag{A.16a}$$

$$\langle\bar{f}_\alpha^{(0)}|=\sum_\beta \bar{c}_{\alpha\beta}\langle\chi_\beta|, \tag{A.16b}$$

where we have dropped the explicit y dependence of the $|f_\alpha^0\rangle$ (and $\langle\bar{f}_\alpha^0|$).

We thus have to solve the 5×5 matrix problem

$$\sum_\beta \epsilon_{\alpha'\beta}c_{\alpha\beta}=\bar{\Lambda}^{(1)}_\alpha(y)c_{\alpha'\beta}, \tag{A.17}$$

where

$$\epsilon_{\alpha'\beta}=\langle\bar{\chi}_{\alpha'}|\Psi^{(1)}(y)|\chi_\beta\rangle. \tag{A.18}$$

In principle, we get contributions to $\epsilon_{\alpha\beta}$ coming from the three categories of terms of order γ discussed in Sec. V.

(i) *Category* (1): This corresponds to pure short range terms. They have been calculated in I and turn out to be all vanishing except

$$[\epsilon_{12}]^{[1]}=-y\sqrt{kT},\quad [\epsilon_{21}]^{[1]}=(-y/\sqrt{kT})\left(\frac{\partial p}{\partial n}\right)^s_T,$$

$$[\epsilon_{25}]^{[1]}=-y\sqrt{2kT/3}\,\frac{1}{nk}\left(\frac{\partial p}{\partial T}\right)^s_n,$$

$$[\epsilon_{52}]^{[1]}=-y\sqrt{2kT/3}\,\frac{1}{nk}\left(\frac{\partial p}{\partial T}\right)^s_n\frac{3k}{2C_v}. \tag{A.19}$$

(ii) *Category* (2): These terms are easily disposed of. Indeed, to first order in γ, the general graph of this type is shown in Fig. 16a. When computing matrix elements of this operator acting on a state $|\chi_\alpha\rangle$ (*or* $\langle\bar{\chi}_\alpha|$), we recover a situation analogous to the one found in the discussion of Fig. 10 and 11, Sec. III. The short range part (the dashed structure in these figures) acting either on $|\chi_\alpha\rangle$ on its right *or* on $\langle\bar{\chi}_\alpha|$ on its left gives zero because of conservation of particles, momentum, and energy. Thus these terms give no contribution to $\epsilon_{\alpha\beta}$:

$$[\epsilon_{\alpha\beta}]^{[2]}=0. \tag{A.20}$$

Similarly, for this same reason, the graph of Fig. 16b will give zero when it is sandwiched between two states $\langle\bar{\chi}_\alpha|$ *and* $|\chi_\alpha\rangle$. This has for a consequence that the graphs of category (2) will not contribute at all to the hydrodynamical modes to order γ^2 and can thus be neglected completely. Incidentally, it is for this same reason that we have already ignored in Sec. IV the contributions of Category (4).

(iii) *Category* (3): Instead of (A.18), let us consider here the slightly more general matrix element

$$\epsilon_{\alpha,\Phi}=\langle\alpha|\Psi^{(1)}(y)|\Phi\rangle, \tag{A.21}$$

where $|\Phi\rangle$ is an arbitrary one particle state.

We write, using (III.14), (V.19), and (V.21),

$$[\epsilon_{\alpha,\Phi}]^{[3]}=[\epsilon_{\alpha,\Phi}]_a+[\epsilon_{\alpha,\Phi}]_b, \tag{A.22}$$

where

$$[\epsilon_{\alpha,\Phi}]_a=\langle\bar{\chi}_\alpha|\frac{-ny}{\sqrt{kT}}V_y^L|2\rangle\langle1|\Phi\rangle, \tag{A.23a}$$

$$[\epsilon_{\alpha,\Phi}]_b=\langle\bar{\chi}_\alpha|\frac{-ny}{\sqrt{kT}}V_y^L\int dv^{N-1}\sum_{k'\neq0}\mathscr{D}^s_{0,k'}v_{ix}\rho^{eq}_{k'}\rangle\langle1|\Phi\rangle. \tag{A.23b}$$

Let us first consider (A.23b). Because the ket $|2\rangle$ is a vector along x, symmetry shows that the only bra $\langle\bar{\chi}_\alpha|$ which might possibly lead to a nonvanishing contribution is $\langle\bar{\chi}_2|\equiv\langle2|$. A simple calculation, based on the symmetrical role played by all the particles and on momentum conservation [see I, Eqs. (A.2), (A.8), (A.9) for a similar argument], then shows that for all α we have

$$[\epsilon_{\alpha,\Phi}]_b=0. \tag{A.24}$$

We are thus left with the simple contribution (A.23a). For the case of interest here, we have then finally

$$[\epsilon_{\alpha,\beta}]^{[3]}=-(ny/\sqrt{kT})V_y^L\delta^{Kr}_{\alpha,2}\delta^{Kr}_{\beta,1}. \tag{A.25}$$

Adding (A.19) and (A.25) and inserting the result into (A.18), the diagonalization of (A.17) is readily achieved. One finds

$$\bar{\Lambda}^{(1)}_{1,2}(y)=\pm iyc(y),\quad \bar{\Lambda}^{(1)}_{3,4,5}(y)=0 \tag{A.26}$$

and the corresponding eigenfunctions are precisely $|f_\alpha^{(0)}\rangle$ and $\langle\bar{f}_\alpha^{(0)}|$ defined by Eq. (V.27), (V.28), where, as usual, we have $k=\gamma y$; moreover, the orthogonality property (V.32) is an immediate consequence of (A.14).

(3) *Order* γ^2: Once the correct basis has been obtained, the second order calculation proceeds in the standard way. From

$$|f_\alpha^{(1)}\rangle=\frac{1}{i(\Psi^{(0)}+i\epsilon)}[-i\Psi^{(1)}(y)+\Lambda^{(1)}_\alpha(y)(1+\dot{\Psi}^{(0)})]|f_\alpha^{(0)}\rangle \tag{A.27}$$

we obtain immediately from (A.5)

$$\bar{\Lambda}^{(2)}_\alpha(y)=\bar{\Lambda}'^{(2)}_\alpha(y)+\bar{\Lambda}''^{(2)}_\alpha(y), \tag{A.28}$$

with

$$\bar{\Lambda}'^{(2)}_\alpha(y)=\lim_{\epsilon\to0}\langle\bar{f}_\alpha^{(0)}|\Psi^{(1)}(y)[i(\Psi^{(0)}+i\epsilon)]^{-1}\times[\Psi^{(1)}(y)-(1+\dot{\Psi}^{(0)})\Lambda^{(1)}_\alpha(y)]|f_\alpha^{(0)}\rangle \tag{A.29a}$$

and

$$\bar{\Lambda}''^{(2)}_\alpha(y)=\lim_{\epsilon\to0}\langle\bar{f}_\alpha^{(0)}|\tfrac{1}{2}\Psi^{(2)}(y)-\ddot{\Psi}^{(0)}(y)[\Lambda^{(1)}_\alpha(y)]^2+i\dot{\Psi}^{(1)}(y)\Lambda^{(1)}_\alpha(y)+i(1+\dot{\Psi}^{(0)})[i(\Psi^{(0)}+i\epsilon)]^{-1}\times\Psi^{(1)}(y)|f_\alpha^{(0)}\rangle. \tag{A.29b}$$

If we take $V_y^L \equiv 0$, i.e., go back to the problem of purely short-range forces, it is readily shown that (A.29) are identical to the eigenvalues defined in I. The only important difference is that we have here only fully connected contributions to Ψ_0^l and its various derivatives, a procedure which was not followed in I; as mentioned in Sec. II, this modification does not affect the final results. Note also that in order to show the equivalence of (A.29b), with I, Eq. (129), some formal manipulations sketched in III, Appendix A, have to be used; we shall not reproduce this calculation here.

It is of course not immediately obvious that the second order eigenvalues (A.28), (A.29) are identical with the k^2 term of (I.3); indeed, in Eqs. (A.29) we have the various derivatives of Ψ with respect to z and γ which are y dependent and we have thus to show that this whole y dependence can be absorbed in the thermodynamic coefficients of (I.3). This calculation has been done for the five eigenvalues $\bar{\Lambda}_\alpha^\gamma(y)$, but it is long and tedious. To keep this appendix within reasonable length, we shall merely illustrate our proof on one nontrivial example, namely the prime part of $\bar{\Lambda}_5^{(2)}(y)$, which describes thermal diffusion.

Using (V.27), (V.28), we obtain easily from (A.29a)

$$\bar{\Lambda}_5'(y) = -[y^2/nC_p(y)]\bar{\kappa}'(y), \tag{A.30}$$

where $C_p(y)$ is defined by (V.4), while

$$\bar{\kappa}'(y) = -\frac{3kn}{2}\lim_{\epsilon\to 0}\langle 5|\Psi^{(1)}(y)[i(\Psi^{(0)}+i\epsilon)]^{-1}\Psi^{(1)}(y) \times\left(|5\rangle + (2/3)^{1/2}\frac{T}{n}\frac{(\partial p/\partial T)_n^s}{(\partial p/\partial n)_T^s + nV_y^L}|1\rangle\right). \tag{A.31}$$

In the development of the first order calculation, we have seen that the only contributions to $\Psi^{(1)}(y)$ which involve long-range forces are of Category (3). Then using (A.23) and (A.24), we rewrite (A.31) as

$$\bar{\kappa}'(y) = -\frac{3kn}{2}\lim_{\epsilon\to 0}\langle 5|\Psi^{(1)}[i(\Psi^{(0)}+i\epsilon)]^{-1}\left(\Psi^{(1)}|5\rangle + \{\Psi^{(1)} + [\Psi^{(1)}(y)]^{[3]}\}(2/3)^{1/2}\frac{T}{n}\frac{(\partial p/\partial T)_n^s}{(\partial p/\partial n)_T^s + nV_y^L}|1\rangle\right), \tag{A.32}$$

where $\Psi^{(1)}$ denotes the pure short-range, y dependent, contribution. Notice that in writing (A.32), we have used the formula

$$[\Psi^{(1)}(y)]^{[3]}|5\rangle = 0, \tag{A.33}$$

an immediate consequence of (V.19).

Substracting from (A.32) the purely short-range thermal conductivity $\bar{\kappa}'^s$ {obtained by formally setting $[\Psi^{(1)}(y)]^{[3]} = 0$, $V_y^L = 0$}, we have thus to prove that

$$\delta\bar{\kappa}'(y) = \bar{\kappa}'(y) - \kappa'^s = -\frac{3kT}{2}\sqrt{2/3}\,\frac{T}{n}\left(\frac{\partial p}{\partial T}\right)_n^s \lim_{\epsilon\to 0}\langle 5|\Psi^{(1)} \times[i(\Psi^{(0)}+i\epsilon)]^{-1}\left[\frac{[\Psi^{(1)}(y)]^{[3]}}{(\partial p/\partial n)_T^s + nV_y^L}|1\rangle + \Psi^{(1)} \times\left(\frac{1}{(\partial p/\partial n)_T^s + nV_y^L} - \frac{1}{(\partial p/\partial n)_T^s}\right)|1\rangle\right] \tag{A.34}$$

vanishes.

A sufficient condition for this property to hold is simply

$$\left(\frac{\partial p}{\partial n}\right)_T^s [\Psi^{(1)}]^{[3]}|1\rangle = nV_y^L\Psi^{(1)}|1\rangle. \tag{A.35}$$

From (V.19)—(V.21), the left-hand side of this equation is simply

$$-ny\beta V_y^L\left(\frac{\partial p}{\partial n}\right)_T^s\left(v_{1x}\varphi^{eq}(v_1) + \sum_{k'\neq 0}\sum_{i=1}^{N}\int dv^{N-1}\mathscr{D}_{0,k'}^s v_{ix}\rho_{k'}^{eq}\right). \tag{A.36}$$

Moreover, we have shown in III, Eq. (III.34a),[31] that

$$\Psi^{(1)}|1\rangle = -\left(v_{1x}\varphi^{eq}(v_1) + \sum_{k'\neq 0}\sum_{i=1}^{N}\int dv^{N-1}\mathscr{D}_{0,k'}^s v_{ix}\rho_{k'}^{eq}\right) \times\beta\left(\frac{\partial p}{\partial n}\right)_T y. \tag{A.37}$$

Inserting (A.36) and (A.37) into (A.35), we see that indeed

$$\delta\bar{\kappa}'(y) = 0, \tag{A.38}$$

which insures that the transport coefficient which appears in $\bar{\Lambda}_5'^{(2)}$ is indeed the pure short-range one; the other cases can be treated similarly.

APPENDIX B: LONG TIME BEHAVIOR OF $X_k(v;\tau)$

We indicate here how (III.10) can be derived from (III.9).

In agreement with our previous hypothesis, we assume that the only zeros of the operator $(z+\Psi_\gamma(y;z))$ which have a vanishing imaginary part when $\gamma\to 0$ are the hydrodynamical eigenvalues $\Lambda_\alpha^\gamma(y)$ [$\alpha\in(G)$]. The analyticity of $\Psi_\gamma(y;z)$ (see, however, footnote 28) allows us then to replace the contour C in Eq. (III.9) by the rectangle R encircling the region

$$-\alpha'\gamma^{1-\delta} \leq \mathrm{Re}\,z \leq \alpha'\gamma^{1-\delta},$$
$$0 \geq \mathrm{Im}\,z \geq -\beta'\gamma^{2-\delta}, \tag{B.1}$$

where α', β', and $\delta < 1$ are real positive numbers. Indeed, in the $\gamma\to 0$ limit, the hydrodynamical poles $z=\Lambda_\alpha^\gamma(y)$ remain inside the rectangle R, while all other singularities of the integrand $\tilde{X}_k(v;z)$ in Eq. (III.9) give contributions at least of the order $\exp-[(\gamma^2\tau)\gamma^{-\delta}]$, which vanishes in the limit $\gamma\to 0$, $(\gamma^2\tau)$ finite, hereafter denoted by $\lim_\gamma$.

We now use the completeness relation [see (A.15)]

$$1 = \sum_n (1+\dot{\Psi}^{(0)})|f_n^{(0)}\rangle\langle\bar{f}_n^{(0)}| = \sum_n |f_n^{(0)}\rangle\langle\bar{f}_n^{(0)}|(1+\dot{\Psi}^{(0)}), \tag{B.2}$$

where

$$|f_n^{(0)}\rangle \equiv |\chi_n\rangle \quad [n\notin(G)]. \tag{B.3}$$

With an obvious notation for the left-hand side, see (II.9), we get then

$$\lim_\gamma X_\gamma(v;\tau) = \lim_\gamma \frac{-1}{2\pi i}\oint_R dz\exp(-iz\tau)\sum_{nn'}|f_n^{(0)}\rangle G_{nn'}\langle\bar{f}_{n'}^{(0)}|, \tag{B.4}$$

where the matrix elements $G_{nn'}$ are defined by

$$G_{nn'} = \langle\bar{f}_n^{(0)}|(1+\dot{\Psi}^{(0)})A^{-1}(1+\dot{\Psi}^{(0)})|f_{n'}^{(0)}\rangle \tag{B.5}$$

and

$$A = z(1+\dot{\Psi}_\gamma(y)) + \Psi_\gamma(y) + (z^2/2)\ddot{\Psi}_\gamma(y), \tag{B.6}$$

a result valid up to order γ^2 when z is restricted by (B.1).

Let us first consider $G_{nn'}$ when $n' \equiv \alpha \in (G)$. We put

$$B = z(1+\dot{\Psi}_\gamma(y)) + \Psi_\gamma(y) - [(\bar{\Lambda}^\gamma_\alpha)^2/2]\ddot{\Psi}_\gamma(y) \tag{B.7}$$

and we use the familiar identity

$$A^{-1} = B^{-1} + A^{-1}(B-A)B^{-1}. \tag{B.8}$$

With the help of (B.2), we obtain then

$$G_{n\alpha} = \sum_{\text{all } n''} \left(\delta^{Kr}_{n'',n} + \sum_{\text{all } n'''} G_{nn'''} g_{n'''n''} \right) H_{n''\alpha}. \tag{B.9}$$

In this formula, we have introduced the matrix elements $g_{nn'}$ and $H_{n''\alpha}$. We have

$$g_{nn'} = -\tfrac{1}{2}(z^2 + \bar{\Lambda}^\gamma_\alpha(y)^2)\langle \bar{f}^{(0)}_n | \ddot{\Psi}_\gamma(y) | f^{(0)}_{n''} \rangle \tag{B.10}$$

$$= 0(\gamma^{2(1-\delta)}), \tag{B.11}$$

this upper bound being a consequence of (B.1) and of the fact that $\bar{\Lambda}^\gamma_\alpha$ is at least of order γ. We also have

$$H_{n''\alpha} = \langle \bar{f}^{(0)}_{n''} | (1+\dot{\Psi}^{(0)}) B^{-1} (1+\dot{\Psi}^{(0)}) | f^{(0)}_\alpha \rangle. \tag{B.12}$$

We then notice that the eigenvalue problem (A.1) leads immediately to

$$B^{-1}(1+\dot{\Psi}_\gamma(y)) | f^\gamma_\alpha \rangle = (z - i\bar{\Lambda}^\gamma_\alpha(y))^{-1} | f^\gamma_\alpha \rangle, \tag{B.13}$$

and, expanding *the numerator* of the left-hand side in power of γ, we get

$$B^{-1}(1+\dot{\Psi}^{(0)}) | f^{(0)}_\alpha \rangle = (z - i\bar{\Lambda}^\gamma_\alpha(y))^{-1} | f^{(0)}_\alpha \rangle (1 + O(\gamma)). \tag{B.14}$$

From the othonormality (A.14), we have thus

$$H_{n''\alpha} = (z - i\bar{\Lambda}^\gamma_\alpha(y))^{-1}(\delta^{Kr}_{n''\alpha} + O(\gamma)). \tag{B.15}$$

Equation (B.9) can thus be simplified to

$$G_{n\alpha} = \left(\delta^{Kr}_{n,\alpha} + \sum_{\text{all } n''} G_{nn''} g_{n''\alpha} \right) (z - i\bar{\Lambda}^\gamma_\alpha(y))^{-1}. \tag{B.16}$$

Notice that (B.16) is not a closed equation because we have to sum over both $n'' \in (G)$ and $n'' \notin (G)$. However, we can now also treat $G_{nn''}$ with $n'' \notin (G)$ by a similar method. Expanding around the zeroth order problem (A.6), we have proved that

$$G_{nn'} = \left(\delta^{Kr}_{n,n'} + \sum_{\text{all } n''} G_{nn'} h_{n''n'} \right) (z - i\mu^0_{n'})^{-1}, \quad (n' \notin (G)) \tag{B.17}$$

where we have for $h_{n''n'}$ (which will not be explicitly needed)

$$h_{n''n'} = O(\gamma^{1-\delta}). \tag{B.18}$$

From (B.11) and (B.18), we see that a self-consistent solution of (B.16), (B.17) is

$$G_{n\alpha} = (z - i\bar{\Lambda}^\gamma_\alpha(y))^{-1}[\delta^{Kr}_{n,\alpha} + O(\gamma^{1-\delta})], \tag{B.19}$$

$$G_{nn'} = (z - i\mu^0_{n'})^{-1}[\delta^{Kr}_{n,n'} + O(\gamma^{1-\delta})] \quad [n' \notin (G)]. \tag{B.20}$$

When (B.19) and (B.20) are inserted into (B.4), and the $\lim_\gamma$ is taken, we obtain indeed Eq. (III.10) by a trivial residue calculation.

Using this result, it is not difficult to show that the dominant contribution involving n intermediate hydrodynamical lines is given by the obvious generalization of (III.15), which corresponds to $n=2$.

APPENDIX C

It is a well-known fact[32] that the Green–Kubo formula (II.5) can be equivalently written as

$$X = \lim_{t\to\infty} \lim_{\infty} \frac{\beta}{\Omega} \int_0^t d\tau \int dr\, dv (J^X - \delta\tilde{J}^X) \exp(-iL\tau) \times (J^X - \delta J^X)\rho^{eq}, \tag{C.1}$$

where $\delta\tilde{J}^X = \sum_i \delta\tilde{J}^X_i$ is an operator, otherwise arbitrary, which satisfies the requirements

$$L\delta\tilde{J}^X = 0, \tag{C.2}$$

$$\int dr\, dv\, \delta\tilde{J}^X (J^X - \delta J^X)\rho^{eq} = 0. \tag{C.3}$$

Usually the choice $\delta\tilde{J}^X \equiv \delta J^X$ is made in order to put Eq. (C.1) in a symmetrical form but it is important to realize that nothing forces us to this provided Eqs. (C.2)–(C.3) are satisfied.

Instead of (II.28), we now get from (C.1)

$$X' = \lim_{t\to\infty} \frac{\beta n}{2\pi} \oint_c \frac{\exp(-izt)}{z} dz \int dv_1 [(0|(J^X_1 - \delta\tilde{J}^X_1)|0) \times \sum_{k\neq 0} (0|(J^X_1 - \delta\tilde{J}^X_1)|k)\mathscr{C}^1_{k0}(z)](1/(z + \Psi^0_i(z)) \times \{[\int dv^{N-1}(0|(J^X - \delta J^X)\rho^{eq}\Omega^N(0)] + \mathscr{D}^1_0(z)\}, \tag{C.4}$$

where we have assumed that the time integral may be performed before the z integration. A similar formula holds for X'' but it will not be needed here.

As (C.4) is the long time limit of a Laplace transform, we may limit ourselves to consider the integrand in the limit of a small z; in particular, using the complete set defined by (A.6), (A.7), and (A.15), we write [we use the same notation as in the purely short-range problem (A.6), although, from the context, it is clear that we are dealing here with the general problem]

$$\frac{1}{z + \Psi^1_0(z)} \approx \frac{1}{z(1+\dot{\Psi}^1_0(i\epsilon)) + \Psi^1_0(i\epsilon) + z^2\ddot{\Psi}^1_0(i\epsilon)/2! + \cdots}$$

$$\approx \frac{1}{z(1+\dot{\Psi}^1_0(i\epsilon)) + \Psi^1_0(i\epsilon)} + \frac{1}{z(1+\dot{\Psi}^1_0(i\epsilon)) + \Psi^1_0(i\epsilon)} \times \frac{z^2}{2!}\ddot{\Psi}^1_0(i\epsilon)\frac{1}{z(1+\dot{\Psi}^1_0(i\epsilon)) + \Psi^1_0(i\epsilon)} + \cdots$$

$$= \sum_{\text{all } m} |\chi_m\rangle(z - i\mu^0_m)^{-1}\langle\bar{\chi}_m| + \sum_{\text{all } m,m'} |\chi_m\rangle \times (z - i\mu^0_m)^{-1}\frac{z^2}{2!}\langle\bar{\chi}_m|\ddot{\Psi}^1_0(i\epsilon)|\chi_{m'}\rangle(z - i\mu^0_{m'})^{-1}\langle\bar{\chi}_{m'}| + \cdots. \tag{C.5}$$

Inserting (C.5) into (C.4), we immediately get two categories of terms in the limit $t \to \infty$:

(1) Those coming from $m \notin (G)$ in the first term of (C.5); they lead to a first order pole at $z=0$ which is readily evaluated:

$$X'_1 = \beta n \sum_{m\notin(G)} \int dv_1 \Big((0|(J^x_1 - \delta\tilde{J}^x_1)|0) + \sum_{k\neq 0} (0|(J^x_1 - \delta\tilde{J}^x_1)|k)\mathscr{C}^1_{k0}(i\epsilon) \Big) |\chi_m\rangle(1/|\mu^0_m|) \times \langle\bar{\chi}_m| \{[\int dv^{N-1}(0|(J^x - \delta J^x)\rho^{eq}\Omega^N|0)] + \mathscr{D}^1_0(i\epsilon)\}. \tag{C.6}$$

(2) Those coming from $m \notin (G)$ [and $m' \in (G)$] in the first (and second) term of (C. 5); before writing these terms explicitly, let us first point out that, using the method of Ref. 13, Appendix 2, it is readily shown that the usual choice[32] of δJ^X guarantees that one has identically

$$\langle \bar{\chi}_\alpha | \{ [\int dv^{N-1} (0 | (J^X - \delta J^X) \rho^{eq} \Omega^N | 0)] + \mathscr{D}_0^l(i\epsilon) \} = 0. \quad (C.7)$$

Taking this result into account, we get simply

$$X_2' = -i\beta n \sum_{\alpha \in (G)} \int dv_1 \Big((0 | (J_1^X - \delta \tilde{J}_1^X) | 0) + \sum_{k \neq 0} (0 | (J_1^X - \delta \tilde{J}_1^X) | k) \mathscr{C}_{k0}^l(i\epsilon) \Big) | \chi_\alpha \rangle \langle \bar{\chi}_\alpha | \frac{\partial \mathscr{D}_0^l(z)}{\partial z} \Big|_{z = i\epsilon} \quad (C.8)$$

It would be fairly inconvenient to retain such a term in the further development of the theory. We shall thus choose $\delta \tilde{J}_1^X$ in such a way that

$$\int dv_1 \Big((0 | (J_1^X - \delta \tilde{J}_1^X) | 0) + \sum_{k \neq 0} (0 | (J_1^X - \delta \tilde{J}_1^X) | k) \mathscr{C}_{k0}^l(i\epsilon) \Big) | \chi_\alpha \rangle = 0 \quad [\alpha \in (G)]. \quad (C.9)$$

Then we simply have

$$X' = X_1' \quad (C.9')$$

and, using (C. 5) in the limit $z \to 0$, we indeed obtain Eq. (II. 38) of the text.

In order to fix $\delta \tilde{J}_1^X$ from the orthogonality condition (C. 9), let us first consider the quantity A_α^X defined by

$$A_\alpha^X = \int dv_1 \Big((0 | J_1^X | 0) + \sum_{k \neq 0} (0 | J_1^X | k) \mathscr{C}_{k0}^l(i\epsilon) \Big) | \chi_\alpha \rangle. \quad (C.10)$$

In the case of thermal conductivity $X = T\kappa$, the method of Ref. 13, Appendix 2, can again be applied and leads to

$$A_\alpha^{T\kappa} = A_2^{T\kappa} \delta_{\alpha,2}^{Kr} \quad (C.11)$$

with

$$A_2^{T\kappa} = (h/n) kT. \quad (C.12)$$

Clearly, the choice $\delta \tilde{J}_1^X$ given in the text by Eq. (II. 40) insures that (C. 9) indeed vanishes; moreover, Eq. (C. 2) and (C. 3) are also trivially satisfied.

Similarly, for $B = 4\eta/3 + \xi$, where

$$J^B = v_{1x}^2 - \frac{1}{2} \sum_j \frac{\partial V}{\partial r_{1j,x}} r_{1j,x},$$

we find

$$A_\alpha^B = A_1^B \delta_{\alpha,1}^{Kr} + A_5^B \delta_{\alpha,5}^{Kr} \quad (C.13)$$

with[33]

$$A_1 = \left[\left(\frac{\partial p}{\partial n} \right)_T - \frac{nV_0}{2} \right], \quad (C.14)$$

$$A_5 = \frac{T}{n} \left(\frac{\partial p}{\partial T} \right)_n. \quad (C.15)$$

A simple calculation then shows that Eqs. (C. 2), (C. 3), and (C. 9) will again be fulfilled provided we take

$$\delta \tilde{J}_1^B = \left(\frac{v_1^2}{2} - \frac{1}{2} \sum_j V(r_{1j}) \right) \left(\frac{\partial p}{\partial e} \right)_n + \left(\frac{\partial p}{\partial n} \right)_e - \frac{nV_0}{2} \left[1 - \left(\frac{\partial p}{\partial e} \right)_n \right]. \quad (C.16)$$

It should be pointed out that $\delta \tilde{J}_1^B$ is *different* from the counterterm appearing at the right of (C. 1); in this case, we have[32]

$$\delta J_1^B = \left[\left(\frac{v_1^2}{2} - \frac{1}{2} \sum_j V(r_{1j}) \right) - \left(\frac{e}{n} \right) \right] \left(\frac{\partial p}{\partial e} \right)_n + \left(\frac{p}{n} \right). \quad (C.17)$$

The reader has remarked that δJ_1^B and $\delta \tilde{J}_1^B$ only differ by constant terms (i.e., independent of r and v); thus, using the conservation of particle number together with (C. 7), we could as well take δJ_1^B as the counterterm at the left of Eq. (II. 38) *provided the* $\epsilon \to 0$ *limit is carefully taken.* This would of course give a more symmetrical form to our basic equations but the quantity appearing at the left of Eq. (II. 38), which would now read

$$\int dv_1 \Big((0 | (J_1^B - \delta J_1^B) | 0) + \sum_{k \neq 0} (0 | (J_1^B - \delta J_1^B) | k) \mathscr{C}_{k0}^l(i\epsilon) \Big), \quad (C.18)$$

would no longer be orthogonal to the invariants of the collision operator [$|\chi_\alpha\rangle, \alpha \in (G)$], [compare with (C. 9)]. As we shall see in the next paper of this series, this feature is not very convenient for explicit calculations and we prefer thus to take $\delta \tilde{J}_1^B \neq \delta J_1^B$ as defined above.

*Permanent address: Institute of Theoretical Physics, University of Warsaw, Poland.

[1] M. Kac, G. Uhlenbeck, P. Hemmer, J. Math. Phys. **4**, 216 (1963); J. Math. Phys. **4**, 229 (1963); J. Math. Phys. **5**, 60 (1964).
[2] N. van Kampen, Phys. Rev. **135**, 362 (1964).
[3] J. Lebowitz and O. Penrose, J. Math. Phys. **7**, 98 (1966).
[4] P. C. Hemmer, J. Math. Phys. **5**, 75 (1964).
[5] J. Lebowitz, G. Stell, and S. Baer, J. Math. Phys. **6**, 1282 (1965).
[6] J. Jalickee, A. Siegert, and D. Vezzetti, J. Math. Phys. **10**, 1442 (1969); J. Math. Phys. **11**, 3168 (1970).
[7] R. Brout, Phys. Rev. **118**, 1009 (1960).
[8] See, for example, K. Kawasaki, Prog. Theor. Phys. **41**, 1190 (1969). J. Deutsch and R. Zwanzig, J. Chem. Phys. **46**, 1612 (1967). R. Montain and R. Zwanzig, J. Chem. Phys. **48**, 1451 (1968).
[9] I. Prigogine, *Non Equilibrium Statistical Mechanics* (Wiley-Interscience, New York, 1962).
[10] P. Resibois, "A perturbative approach to irreversible statistical mechanics," in *Physics of Many-Particle Systems*, edited by E. Meeron (Gordon and Breach, New York, 1966).
[11] P. Résibois, J. Chem. Phys. **41**, 2979 (1964).
[12] J. Piasecki, Physica **58**, 360 (1972).
[13] P. Résibois, J. Stat. Phys. **2**, 21 (1970), denoted by I in Appendix A.
[14] We formally set the mass $m = 1$. Moreover, when no ambiguity is possible, we do not write explicitly the vector symbols.
[15] The factor Ω^N which always appears in combination with matrix elements of ρ^{eq} comes from our definition (II.9) of Fourier transforms. It insures that the one-particle quantities $\Psi_0^l, \mathscr{D}_0^l, \mathscr{C}_0^l, \mathscr{F}_0^l$ have a finite thermodynamic limit.
[16] We always read a graph from right to left.
[17] See, for example, H. Mori, Prog. Theor. Phys. **28**, 763 (1962).
[18] In this part of the discussion, it is convenient to use the time-dependent version of the theory. See. P. Resibois and M. de Leener, Phys. Rev. **178**, 806 (1969).
[19] Although we generally do not make vector symbols explicit, we write down explicitly d^3k for the volume element in wave vector space to keep easy track of the k-dimensionality of the various expressions to be discussed.
[20] R. Elliot and L. de Sobrino, Can. J. Phys. **45**, 3555 (1967).
[21] Although we have here a graphical method for a classical problem, we find it convenient to use the language of quantum field theory. See, for example, A. Abrikosov, L. Gorkov, and I. Dzyaloshinsky, *Methods of Quantum Field Theory in Statistical Physics* (Prentice-Hall, Englewood Cliffs, N. J., 1965).
[22] We use here the abstract vector space notation for one-particle velocity functions, defined in Ref. 13.

[23]Note that for this argument to be valid, it is essential to have first performed the renormalization of each single line.

[24]Let us stress that, in writing Eq. (V. 13), we have used (III.10) which is precisely what we want to prove! The calculation of this section will thus give us a self-consistent proof of our conjectures of Sec. III.

[25]The self-diffusion coefficient does not enter into the present analysis because it requires the consideration of the self-diffusion mode, and from this point of view, is similar to a fluid mixture problem.

[26]In this appendix, we take k along the x axis and we can thus neglect the dependence of the eigenfunctions on 1_k.

[27]P. Resibois, Bull. Cl. Sci. Acad. R. Belg. **56**, 160 (1970), denoted by II in Appendix A.

[28]P. Resibois, Physica (Utr.) **49**, 591 (1971), denoted by III in Appendix A.

[29]It has been recently discovered that, even with purely short-range forces, the operator $\Psi^0(z)$ is nonanalytic close to $z = 0$: $\Psi^0(z) - \Psi^0(0) \alpha z^{(1/2)}$. [See R. Dorfman and E. Cohen, Phys. Rev. Lett. **25**, 1257 (1970), for a microscopic analysis of this property in a simple model for fluids.] This difficulty, which is closely connected to the slow decay of the correlation functions for long times, makes the very existence of Ψ^0 doubtful. However, the final result of our present calculation is entirely expressible in terms of thermodynamic properties and of transport coefficients, which, in three dimensions, are assumed to exist. This justifies *a posteriori* the formal procedure followed here. Although we are perfectly aware of this difficulty, which is also implicit in I–III, we shall thus assume that $\Psi^0(z)$ may be *formally* considered as an analytical function of z close to the real axis.

[30]In I, we considered an operator Ω_0^{-1} instead of $(1 + \Psi^{(0)})$. However, it is easily verified that, for the matrix elements needed here, these two operators lead to the same result [see I, Eqs. (A2), (A3)].

[31]We should point out an unfortunate misprint in this latter formula: The factor $(\partial p/\partial T)_n$ should of course be replaced by $(\partial p/\partial n)_T$.

[32]I. MacLennan, *Advances in Chemical Physics* (Interscience, New York, 1963), Chap. V, p. 261.

[33]It is an unfortunate feature of the perturbation method used here that the terms of first order in the potential—involving V_0—often play a special role (see V. 29); of course, this type of term always cancels out in the final physical result.

Matrices of finite Lorentz transformations in a noncompact basis. III. Completeness relation for $O(2,1)$

N. Mukunda*

Tata Institute of Fundamental Research, Bombay, India
(Received 12 January 1972)

A parametrization of the elements of the three-dimensional Lorentz group $O(2, 1)$, suited to the use of a noncompact $O(1, 1)$ basis in its unitary representations, is derived and used to set up the representation matrices for the entire group. The Plancherel formula for $O(2, 1)$ is then expressed in this basis.

INTRODUCTION

In the two previous papers of this series we have dealt with the problem of setting up the unitary irreducible representations (UIR's) of the group $O(2, 1)$ in a basis in which the (noncompact) generator of an $O(1, 1)$ subgroup is diagonal.[1] The representations of physical interest are actually the single-valued UIR's of the spinor group $SU(1, 1)$, which include both single-and double-valued representations of $O(2, 1)$. In Papers I and II we have worked out explicitly the matrices that represent elements on the one-parameter subgroups generated by the two generators of $SU(1, 1)$ other than the one that is diagonal; this was done for all the UIR's of $SU(1, 1)$, namely in the terminology of Bargmann, for the discrete, the continuous nonexceptional, and the continuous exceptional series of UIR's.[2]

In Bargmann's paper the following important theorem was proved: every (Lebesgue) square-integrable function on the group $SU(1, 1)$ can be expanded in terms of the matrix elements of the representation matrices belonging to a subset of the set of all UIR's of $SU(1, 1)$. The integration over $SU(1, 1)$ is the usual left and right invariant one, and the UIR's not needed for the above purpose are those of the continuous exceptional series and the lowest ones, $D^{\pm}_{1/2}$, of the discrete series. This theorem was proved by examining the properties of the representation matrices in a basis in which the compact subgroup $O(2)$ was diagonal. But it is clear that the real content of this theorem is independent of the basis chosen in setting up the representations of $SU(1, 1)$; the purpose of this paper is to transcribe the result of Bargmann to the situation wherein all the UIR's of $SU(1, 1)$ are obtained in a noncompact basis.[3]

In I we examined in detail the properties of two particular families of elements of $SU(1, 1)$; these types were chosen because their representative matrices were especially simple in an $O(1, 1)$ basis, and it was noted that elements of these two types do not collectively exhaust all of $SU(1, 1)$. We shall show in Sec. 1 that almost every element of $SU(1, 1)$ can be uniquely written in one of five distinct canonical forms, these forms being adapted to the use of an $O(1, 1)$ basis in the UIR's; and in every case, the representation matrix in such a basis can be easily related to the results previously established in I and II. The elements of $SU(1, 1)$ that are left out in this description form a set of measure zero. The existence of these five canonical forms is the analog in the present case to the "Euler" type decomposition for $SU(1, 1)$ used by Bargmann; we give simple geometrical arguments that explain this new parametrization of $SU(1, 1)$. At the level of the group $O(2, 1)$, just three canonical forms suffice. In Sec. 2, we state Bargmann's expansion formula in its original form, and then show how it may be rewritten using the representation matrices in the $O(1, 1)$ basis. We adhere throughout to the notations and conventions of I, II and our earlier papers on this subject.[4]

I. PARAMETRIZATION OF THE GROUP SU(1,1)

The defining representation of the group $SU(1, 1)$ is provided by the collection of two-dimensional complex matrices of the form

$$\begin{pmatrix} \alpha & \beta \\ \beta^* & \alpha^* \end{pmatrix}, \qquad |\alpha|^2 - |\beta|^2 = 1. \tag{1.1}$$

A convenient set of real parameters may then be chosen by writing

$$\alpha = \cosh\xi e^{i\phi}, \quad \beta = \sinh\xi e^{i\psi} \tag{1.2}$$

and choosing the ranges of ξ, ϕ, ψ to be

$$0 \leq \xi < \infty, \quad -\pi \leq \phi, \quad \psi \leq \pi. \tag{1.3}$$

We may denote a general element in $SU(1, 1)$ by $g(\xi, \phi, \psi)$. The Lie algebra of $SU(1, 1)$ is spanned by three elements J_0, J_1, J_2 obeying the commutation rules

$$-i[J_0, J_1] = J_2, \quad -i[J_0, J_2] = -J_1, \quad -i[J_1, J_2] = -J_0, \tag{1.4}$$

and the quadratic combination

$$Q = J_1^2 + J_2^2 - J_0^2 \tag{1.5}$$

commutes with all the J's, so it is a Casimir operator which becomes equal to a real number in a UIR. In the defining representation (1.1), we identify the J's as follows:

$$J_0 = \tfrac{1}{2}\sigma_3, \quad J_1 = (i/2)\sigma_2, \quad J_2 = -(i/2)\sigma_1. \tag{1.6}$$

J_0 generates the compact $O(2)$ subgroup of $SU(1, 1)$, while J_1 and J_2 each generate a noncompact $O(1, 1)$ subgroup. In I and II, the UIR's of $SU(1, 1)$ were set up with J_2 diagonal; whenever we refer to a noncompact basis, we shall mean the one consisting of eigenvectors of J_2.

Elements of the following two types

$$h(\zeta, \mu, \zeta') = e^{i\zeta J_2} e^{i\mu J_0} e^{i\zeta' J_2}, \tag{1.7a}$$

$$k_0(\zeta, \nu, \zeta') = e^{i\zeta J_2} e^{i\nu J_1} e^{i\zeta' J_2}, \tag{1.7b}$$

evidently have particularly simple representation matrices in the noncompact basis, since the first and last factors become trivial. [We write $k_0(\zeta, \nu, \zeta')$ now for what was denoted as $k(\zeta, \nu, \zeta')$ in I, for reasons that will become clear shortly.] By allowing the parameters

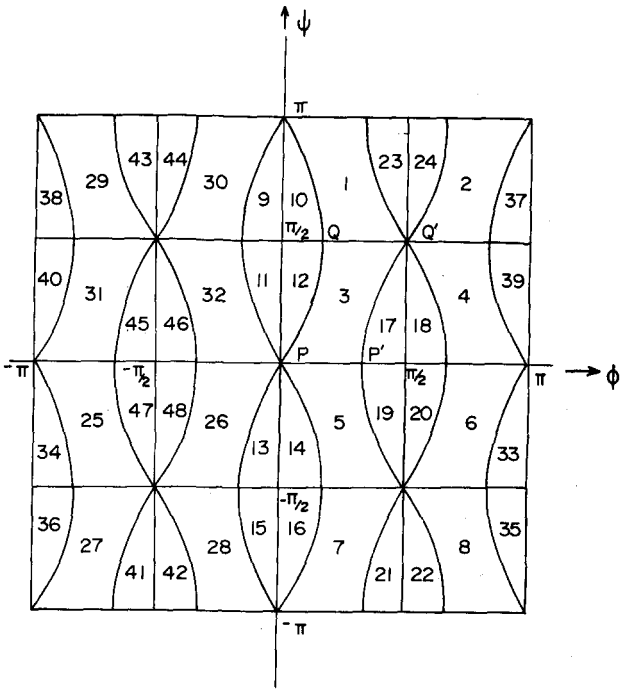


FIG. 1. Various regions in $SU(1,1)$: $R=(1$ to 8, 25 to 32); $S_0=(9$ to 16); $S_1=(17$ to 24); $S_2=(33$ to 40); $S_3=(41$ to 48).

to vary in the region $-\infty<\zeta,\zeta'<\infty$, $-2\pi\leq\mu\leq2\pi$ in the case of $h(\zeta,\mu,\zeta')$, and in the region $-\infty<\zeta,\zeta',\nu<\infty$ in the case of $k_0(\zeta,\nu,\zeta')$, we showed in I that two distinct nonoverlapping regions of $SU(1,1)$ are obtained. Stated in another way, for each possible value of the coordinate ξ appearing in (1.2) and (1.3), only those elements $g(\xi,\phi,\psi)$ for which ϕ,ψ lie in certain restricted regions of the $\phi-\psi$ plane can be rewritten in the form $h(\zeta,\mu,\zeta')$ for some values of ζ, μ, and ζ'; and only if ϕ,ψ lie in certain other restricted regions can $g(\xi,\phi,\psi)$ be put into the form $k_0(\zeta,\nu,\zeta')$ for suitable ζ, ν, and ζ'. Even after considering these two kinds of elements, certain portions of the $\phi-\psi$ plane were left uncovered.

We would like to be able to express (almost) every element $g(\xi,\phi,\psi)$ in a form somewhat like (1.7a, b), with the first and last factors being $O(1,1)$ transformations generated by J_2, and the factor (or factors) in between involving only one parameter. Making use of the identification (1.6) for the generators in the defining representation, as well as the results of I, it turns out that if in addition to $h(\zeta,\mu,\zeta')$ and $k_0(\zeta,\nu,\zeta')$ we consider the following three kinds of elements,

$$k_n(\zeta,\nu,\zeta')=e^{i\zeta J_2}e^{i\nu J_1}e^{in\pi J_0}e^{i\zeta' J_2},$$
$$-\infty<\zeta,\nu,\zeta'<\infty,\quad n=1,2,3, \tag{1.8}$$

then almost the entire $\phi-\psi$ plane is covered. The regions not obtained are the one-dimensional boundaries between the open regions that are covered; these boundaries are of measure zero. Further, the five types of elements $h(\zeta,\mu,\zeta')$ and $k_n(\zeta,\nu,\zeta')$ for $n=0,1,2,3$ are all distinct in that apart from trivial cases an element of one type cannot coincide with an element of another type. In this way we achieve a parametrization of the entire group $SU(1,1)$ suited to the use of the noncompact basis in the UIR's. It is also easily verified that with the ranges we have chosen for ζ, μ, ζ' in the case of $h(\zeta,\mu,\zeta')$ and ζ, ν, ζ' in the cases of $k_n(\zeta,\nu,\zeta')$, two distinct sets of values for these parameters always lead to distinct elements of $SU(1,1)$.

We will refer to the region in $SU(1,1)$ consisting of the elements $h(\zeta,\mu,\zeta')$ as R, and to the four regions consisting of the elements $k_n(\zeta,\nu,\zeta')$ for $n=0,1,2,3$ as S_n. These five regions may be depicted in the $\phi-\psi$ plane as shown in Fig. 1. For each value of ξ, this plane (or more precisely the square $-\pi\leq\phi,\ \psi\leq\pi$) is divided in a particular way into 48 regions. The equations to the arcs PQ and $P'Q'$ within the square $0\leq\phi,\ \psi\leq\pi/2$ suffice to determine, by suitable translations and reflections of these two arcs, the break-up of the square $-\pi\leq\phi,\ \psi\leq\pi$ into 48 pieces:

$$PQ:\ \sin\psi=\sin\phi\coth\xi,$$
$$P'Q':\ \cos\psi=\cos\phi\coth\xi. \tag{1.9}$$

Regions 1 to 8 and 25 to 32 together give the portion R in $SU(1,1)$; 9 to 16 give S_0, 17 to 24 give S_1, 33 to 40 give S_2, and 41 to 48 give S_3. The correspondence between $SU(1,1)$ and $O(2,1)$ is two-to-one. One checks that elements in $SU(1,1)$ in the regions 1 to 8 and those in the regions 25 to 32 correspond to the same region in $O(2,1)$; similarly S_0 and S_2 are mapped into the same portion of $O(2,1)$, and so are S_1 and S_3. Thus almost every element in $O(2,1)$ can be put into one of the three forms $h(\zeta,\mu,\zeta')$, $k_0(\zeta,\nu,\zeta')$, and $k_1(\zeta,\nu,\zeta')$ with the ranges now being $-\infty<\zeta,\zeta',\nu<\infty$, $0\leq\mu\leq2\pi$.

A simple geometrical argument can be given to explain why three canonical forms are needed to represent elements of $O(2,1)$, the requirements on the forms being (a) the first and last factor should be $O(1,1)$ transformations generated by J_2, (b) simplicity. Consider the action of $O(2,1)$ on vectors in three-dimensional space–time, J_0 generating rotations in the x-y plane, J_1 generating pure Lorentz transformations in the x-t variables, J_2 generating similar transformations in the y-t variables. Keeping condition (a) in mind, let us define the three-dimensional spacelike vector $p^{(0)}$ by

$$p^{(0)}=(0,1,0). \tag{1.10}$$

The first, second, and third components are the t, x, and y components, respectively. The elements of $O(2,1)$ leaving $p^{(0)}$ invariant are precisely the $O(1,1)$ transformations with J_2 as generator. By applying all possible $O(2,1)$ transformations to $p^{(0)}$ we obtain all possible three-vectors $p=(E,\ p_x,\ p_y)$ such that

$$p_x^2+p_y^2-E^2=1. \tag{1.11}$$

Let us write A for a general element of $O(2,1)$, and let it take the vector $p^{(0)}$ into $Ap^{(0)}=p$. If for each p, we choose in some way a definite element $L(p)$ in $O(2,1)$ such that

$$L(p)p^{(0)}=p, \tag{1.12}$$

then it is clear that A can differ from $L(Ap^{(0)})$ only by an element of the form $e^{i\zeta J_2}$ on the right:

$$A=L(Ap^{(0)})e^{i\zeta J_2}. \tag{1.13}$$

The form we desire for all A is

$$A=e^{i\zeta J_2}Xe^{i\zeta' J_2}, \tag{1.14}$$

with a suitable set of choices for the element X; com-

paring with (1.13) this means that the $L(p)$ must be chosen for each p in the form

$$L(p)=e^{i\zeta J_2}X(p), \quad \zeta=\zeta(p). \tag{1.15}$$

Given that the last factor in $L(p)$ is also an $O(1,1)$ transformation, it is clear that the element $X(p)$ in $O(2,1)$ must perform the task of taking $p^{(0)}$ into a vector whose x component is p_x and whose t and y components can lead via a further $O(1,1)$ transformation to the desired values E, p_y. We then easily arrive at the following list of distinct configurations for $p=(E,\ p_x,\ p_y)$ and corresponding simple choices for $L(p)$:

(a) $E^2>p_y^2,\ p_x>0: L(p)=\exp(i\zeta J_2)\exp(i\nu J_1)$,

(b) $E^2>p_y^2,\ p_x<0: L(p)=\exp(i\zeta J_2)\exp(i\nu J_1)\exp(i\pi J_0)$,

(c) $E^2<p_y^2: L(p)=\exp(i\zeta J_2)\exp(i\mu J_0)$,

(d) $E=\pm p_y,\ p_x=1: L(p)=\exp[ib(J_1\pm J_0)]$,

(e) $E=\pm p_y,\ p_x=-1: L(p)=\exp[ib(J_1\pm J_0)]\exp(i\pi J_0)$.

(1.16)

Elements A in $O(2,1)$ such that $Ap^{(0)}$ is a vector of the type (d) or (e) form a two-dimensional family, which is a set of measure zero. All other elements of $O(2,1)$ are accounted for by considering the three types $h(\zeta,\mu,\zeta')$, $k_0(\zeta,\nu,\zeta')$, and $k_1(\zeta,\nu,\zeta')$, as asserted earlier. In going back to $SU(1,1)$, all that is involved is a further doubling.[5]

The matrices representing the elements $h(\zeta,\mu,\zeta')$ and $k_0(\zeta,\nu,\zeta')$ in the UIR's of $SU(1,1)$ in a noncompact basis were determined in I and II. We now need an extension of this work to cover the other types of elements $k_n(\zeta,\nu,\zeta')$, $n=1,2,3$. We use the symbol $\mathfrak{R}$ to denote a general UIR of $SU(1,1)$; here we only deal with those UIR's that appear in Bargmann's completeness relation, so $\mathfrak{R}=(k,+)$ for the positive discrete series D_k^+, $\mathfrak{R}=(k,-)$ for the negative discrete series D_k^-, $\mathfrak{R}=(s,0)$ for the continuous integral nonexceptional series, and $\mathfrak{R}=(s,\frac{1}{2})$ for the continuous half-integral series. The ranges for k and s are $k=1,\frac{3}{2},2,\ldots$, $0\le s<\infty$, and $Q=k(1-k)$ or $\frac{1}{4}+s^2$ as the case may be. An element of the noncompact basis in the UIR $\mathfrak{R}$ is written

$$|\mathfrak{R};p,a\rangle, \quad -\infty<p<\infty, \quad a=\pm, \tag{1.17}$$

p is the eigenvalue of J_2, and the multiplicity index a is understood to be absent if $\mathfrak{R}=(k,\pm)$. The choice of these basis vectors for all the UIR's is explained in detail in I and II. We now define the representation matrices for the five types of elements in $SU(1,1)$ in this fashion:

$$\langle\mathfrak{R};p',b|h(\zeta,\mu,\zeta')|\mathfrak{R};p,a\rangle=\exp[i(\zeta p'+\zeta' p)]\mathscr{g}_{ba}^{(\mathfrak{R})}(p',p;\mu), \tag{1.18a}$$

$$\langle\mathfrak{R};p',b|k_n(\zeta,\nu,\zeta')|\mathfrak{R};p,a\rangle=\exp[i(\zeta p'+\zeta' p)]\mathfrak{F}_{ba}^{(\mathfrak{R})}(p',p;\nu;n) \quad n=0,1,2,3. \tag{1.18b}$$

(Here, and in the rest of this paper, it is to be understood that in case $\mathfrak{R}$ denotes a discrete series UIR, the labels a, b are to be dropped.) The functions $\mathscr{g}$ defined in (1.18a) are the same as those defined and evaluated in I and II. The set of functions $\mathfrak{F}_{ba}^{(\mathfrak{R})}(p',p;\nu;0)$, corresponding to $n=0$, are the ones evaluated in I, II and written there as $\mathfrak{F}_{ba}^{(\mathfrak{R})}(p',p;\nu)$. The cases $n=1,2,3$ are immediately related to the case $n=0$. Basically we need to know the action of the unitary operator $\exp(i\pi J_0)$ on the noncompact basis in the various UIR's. This is given in I and II and reads

$$\exp(i\pi J_0)|\mathfrak{R};p,a\rangle=\begin{cases}\exp(i\pi\eta k)|k,\eta;-p\rangle & \text{for } \mathfrak{R}=(k,\eta),\ \eta=\pm,\\ a|s,0;-p,a\rangle & \text{for } \mathfrak{R}=(s,0),\\ a|s,\frac{1}{2};-p,-a\rangle & \text{for } \mathfrak{R}=(s,\frac{1}{2}).\end{cases} \tag{1.19}$$

We have then the following system of equations relating the cases $n=1,2,3$ to the case $n=0$:

$$\mathfrak{F}_{ba}^{(\mathfrak{R})}(p',p;\nu;0)\equiv\mathfrak{F}_{ba}^{(\mathfrak{R})}(p',p;\nu) \text{ of I, II,} \tag{1.20a}$$

$$\begin{aligned}\mathfrak{F}_{ba}^{(\mathfrak{R})}(p',p;\nu;1)&=\exp(i\pi\eta k)\mathfrak{F}^{(k,\eta)}(p',-p;\nu) \text{ if } \mathfrak{R}=(k,\eta)\\ &=a\mathfrak{F}_{b,a}^{(s,0)}(p',-p;\nu) \text{ if } \mathfrak{R}=(s,0)\\ &=a\mathfrak{F}_{b,-a}^{(s,1/2)}(p',-p;\nu) \text{ if } \mathfrak{R}=(s,\tfrac{1}{2}),\end{aligned} \tag{1.20b}$$

$$\begin{aligned}\mathfrak{F}_{ba}^{(\mathfrak{R})}(p',p;\nu;2)&=(-1)^{2k}\mathfrak{F}^{(k,\eta)}(p',p;\nu) \text{ if } \mathfrak{R}=(k,\eta)\\ &=(-1)^{2\epsilon}\mathfrak{F}_{ba}^{(s,\epsilon)}(p',p;\nu) \text{ if } \mathfrak{R}=(s,\epsilon),\end{aligned} \tag{1.20c}$$

$$\begin{aligned}\mathfrak{F}_{ba}^{(\mathfrak{R})}(p',p;\nu;3)&=(-1)^{2k}\exp(i\pi\eta k)\mathfrak{F}^{(k,\eta)}(p',-p;\nu) \text{ if } \mathfrak{R}=(k,\eta)\\ &=a\mathfrak{F}_{b,a}^{(s,0)}(p',-p;\nu) \text{ if } \mathfrak{R}=(s,0)\\ &=-a\mathfrak{F}_{b,-a}^{(s,1/2)}(p',-p;\nu) \text{ if } \mathfrak{R}=(s,\tfrac{1}{2}).\end{aligned} \tag{1.20d}$$

With this, the task of computing the representation matrices for almost all elements of $SU(1,1)$ in all the UIR's of interest, in the noncompact basis, is done. We will not write down here the explicit expressions for the $\mathfrak{F}$'s and $\mathscr{g}$'s in terms of hypergeometric functions, since all this has been spelt out in detail in I and II. We now use these results to develop the completeness property of these UIR's, in the noncompact basis.

2. THE COMPLETENESS RELATION

We begin by setting up an adequate notation with which to express Bargmann's theorem. Every element g in $SU(1,1)$ can be decomposed in the "Euler" form as follows:

$$g(\alpha,\beta,\gamma)=\exp(-i\alpha J_0)\exp(-i\beta J_2)\exp(-i\gamma J_0). \tag{2.1}$$

The parameters α,β,γ are all real [α and β here must not be confused with the complex quantities appearing in (1.1) and (1.2)]; by choosing the ranges

$$0\le\beta<\infty, \quad 0\le\alpha\le 2\pi, \quad -2\pi\le\gamma\le 2\pi, \tag{2.2}$$

we ensure that except for a set of measure zero every element of $SU(1,1)$ appears just once.[6] If $f(g)$ is a suitable function over $SU(1,1)$, its invariant integral is defined by

$$\int_{SU(1,1)}dgf(g)\equiv\frac{1}{2}\int_0^\infty\sinh\beta d\beta\int_0^{2\pi}\frac{d\alpha}{2\pi}\int_{-2\pi}^{2\pi}\frac{d\gamma}{4\pi}f(\alpha,\beta,\gamma). \tag{2.3}$$

Let us write $\mathscr{D}^{(\mathfrak{R})}(g)$ for the infinite-dimensional unitary matrix that represents the element g in the UIR $\mathfrak{R}$; it may be expressed in any basis. In the basis in which

J_0 is diagonal, with eigenvalues m, m', ..., we have with the decomposition (2.1) for g,

$$\mathfrak{D}^{(\mathfrak{R})}_{m'm}[g(\alpha,\beta,\gamma)] \equiv \langle \mathfrak{R};m' | \exp(-i\alpha J_0)\exp(-i\beta J_2) \exp(-i\gamma J_0) | \mathfrak{R};m\rangle = \exp[-i(\alpha m'+\gamma m)]d^{(\mathfrak{R})}_{m'm}(\beta). \tag{2.4}$$

The expressions for $d^{(\mathfrak{R})}_{m'm}(\beta)$ are given in Bargmann's paper[2]; the five sets of functions $\mathcal{G}, \mathcal{F}$ are what replace these d's in replacing the compact by the noncompact basis.

For a function $\tilde{f}(\mathfrak{R})$ defined over the UIR's $\mathfrak{R}$, we define a summation over the UIR's by the formula

$$\int d\mathfrak{R}\tilde{f}(\mathfrak{R}) \equiv \sum_{k=1,3/2,\ldots}^{\infty} \sum_{\eta=\pm} \tilde{f}(k,\eta) + \sum_{\epsilon=0,1/2} \int_0^\infty ds\ \tilde{f}(s,\epsilon). \tag{2.5}$$

The Kronecker symbol $\delta(\mathfrak{R}',\mathfrak{R})$ that goes with this is

$$\begin{aligned}\delta(\mathfrak{R}',\mathfrak{R}) &= \delta_{k'k}\delta_{\eta'\eta} \text{ if } \mathfrak{R}=(k,\eta),\quad \mathfrak{R}'=(k',\eta'),\\ &= \delta_{\epsilon'\epsilon}\delta(s'-s) \text{ if } \mathfrak{R}=(s,\epsilon),\quad \mathfrak{R}'=(s',\epsilon'),\\ &= 0 \text{ otherwise.}\end{aligned} \tag{2.6}$$

We also need to define a positive weight function $\mu(\mathfrak{R})$ as follows:

$$\begin{aligned}\mu(\mathfrak{R}) &= (2k-1)^{1/2} \text{ if } \mathfrak{R}=(k,\eta),\\ &= (2s/\coth\pi s)^{1/2} \text{ if } \mathfrak{R}=(s,0),\\ &= (2s/\tanh\pi s)^{1/2} \text{ if } \mathfrak{R}=(s,\tfrac{1}{2}).\end{aligned} \tag{2.7}$$

Now let $\mathcal{H}$ be the Hilbert space of all Lebesgue square integrable functions over $SU(1,1)$, that is of all functions $f(g)$ such that

$$(f,f) \equiv \int dg\ |f(g)|^2 < \infty. \tag{2.8}$$

Bargmann's theorem then asserts that such an f can be expanded in the form

$$f(g) = \int d\mathfrak{R}\mu(\mathfrak{R}) \sum_{mn} \tilde{f}_{mn}(\mathfrak{R})\mathfrak{D}^{(\mathfrak{R})}_{mn}(g), \tag{2.9}$$

with the coefficients given by

$$\tilde{f}_{mn}(\mathfrak{R}) = \mu(\mathfrak{R}) \int dg \mathfrak{D}^{(\mathfrak{R})}_{mn}(g)^* f(g), \tag{2.10}$$

and we will then have

$$\int dg\,|f(g)|^2 = \int d\mathfrak{R} \sum_{mn} |\tilde{f}_{mn}(\mathfrak{R})|^2. \tag{2.11}$$

In particular, we have the orthogonality relations for the $\mathfrak{D}$ matrices:

$$\int dg \mathfrak{D}^{(\mathfrak{R})}_{mn}(g)^* \mathfrak{D}^{(\mathfrak{R}')}_{m'n'}(g) = \delta(\mathfrak{R},\mathfrak{R}')\delta_{mm'}\delta_{nn'}/\mu(\mathfrak{R})\mu(\mathfrak{R}'). \tag{2.12}$$

If we now use (2.10) in (2.9) and write the latter in the form

$$f(g) = \int d\mathfrak{R}\mu^2(\mathfrak{R}) \int dg'\, \mathrm{Tr}[\mathfrak{D}^{(\mathfrak{R})}(g'^{-1})\mathfrak{D}^{(\mathfrak{R})}(g)] f(g'), \tag{2.13}$$

we can transcribe this formula to the noncompact basis.

The Hilbert space $\mathcal{H}$ must be given a different description corresponding to the new parametrization of $SU(1,1)$. A function $f(g)$ defined on the group now amounts to a column vector made up of five entries:

$$f(g) \equiv \begin{pmatrix} f(\zeta,\mu,\zeta') \\ \\ f_n(\zeta,\nu,\zeta') \end{pmatrix}, \quad n=0,1,2,3; \tag{2.14}$$

the first entry is just the value of f when the element g lies in the region R, while $f_n(\zeta,\nu,\zeta')$ is its value when g belongs to S_n. The Jacobian to go from the parameters (α,β,γ) in (2.1) to these new ones can be easily worked out in each region, and the invariant integral of (2.3) takes the form

$$\int dg f(g) \equiv (32\pi^2)^{-1} \int_{-\infty}^{\infty} d\zeta' \int_{-\infty}^{\infty} d\zeta \left(\int_{-2\pi}^{2\pi} d\mu\,|\sin\mu|\, f(\zeta,\mu,\zeta') + \sum_{n=0}^{3} \int_{-\infty}^{\infty} d\nu\,|\sinh\nu|\, f_n(\zeta,\nu,\zeta') \right). \tag{2.15}$$

$\mathcal{H}$ can now be described as consisting of those column vectors of the form (2.14) for which

$$(f,f) \equiv (32\pi^2)^{-1} \int_{-\infty}^{\infty} d\zeta' \int_{-\infty}^{\infty} d\zeta \left(\int_{-2\pi}^{2\pi} d\mu\,|\sin\mu|\,|f(\zeta,\mu,\zeta')|^2 + \sum_{n=0}^{3} \int_{-\infty}^{\infty} d\nu\,|\sinh\nu|\,|f_n(\zeta,\nu,\zeta')|^2 \right) < \infty. \tag{2.16}$$

For the elements of $\mathcal{H}$ we derive the new form of the expansion formula by computing the trace that appears in (2.13) in the noncompact basis. As the elements g, g' vary in each of the regions R, S_n of $SU(1,1)$, the matrix elements of $\mathfrak{D}^{(\mathfrak{R})}(g)$ and $\mathfrak{D}^{(\mathfrak{R})}(g'^{-1})$ can be expressed in terms of the $\mathcal{F}$'s and $\mathcal{G}$'s. Thus we obtain, for column vectors that fulfill (2.16), the formula

$$\begin{pmatrix} f(\zeta,\mu,\zeta') \\ \\ f_n(\zeta,\nu,\zeta') \end{pmatrix} = \int d\mathfrak{R}\mu(\mathfrak{R}) \sum_{ab} \int_{-\infty}^{\infty} dp' \int_{-\infty}^{\infty} dp\, \tilde{f}_{p'b;pa}(\mathfrak{R}) \times \exp[i(\zeta p' + \zeta' p)] \begin{pmatrix} \mathcal{G}^{(\mathfrak{R})}_{ba}(p',p;\mu) \\ \\ \mathcal{F}^{(\mathfrak{R})}_{ba}(p',p;\nu;n) \end{pmatrix}, \tag{2.17}$$

the expansion coefficients being determined by

$$\tilde{f}_{p'b;pa}(\mathfrak{R}) = \mu(\mathfrak{R})(32\pi^2)^{-1} \int_{-\infty}^{\infty} d\zeta' \int_{-\infty}^{\infty} d\zeta \exp[-i(\zeta p' + \zeta' p)] \left(\int_{-2\pi}^{2\pi} d\mu\,|\sin\mu|\, \mathcal{G}^{(\mathfrak{R})}_{ba}(p',p;\mu)^* f(\zeta,\mu,\zeta') + \sum_{n=0}^{3} \int_{-\infty}^{\infty} d\nu\,|\sinh\nu|\,(\mathcal{F}^{(\mathfrak{R})}_{ba}(p',p;\nu;n))^* f_n(\zeta,\nu,\zeta') \right) \tag{2.18}$$

The previous Eq. (2.11) is to be replaced by

$$(f,f) = \int d\mathfrak{R} \sum_{ab} \int_{-\infty}^{\infty} dp' \int_{-\infty}^{\infty} dp\,|\tilde{f}_{p'b;pa}(\mathfrak{R})|^2, \tag{2.19}$$

the left-hand side being defined in (2.16). The above three equations constitute the statement of Bargmann's theorem in an $O(1,1)$ basis. If as a particular case we consider a function $f(g)$ which is actually defined on $O(2,1)$, then the entries in the column vector (2.14) obey

$$\begin{aligned} &f(\zeta,\mu,\zeta') = f(\zeta,\mu\pm 2\pi,\zeta'), \quad f_0(\zeta,\nu,\zeta') = f_2(\zeta,\nu,\zeta'),\\ &f_1(\zeta,\nu,\zeta') = f_3(\zeta,\nu,\zeta'). \end{aligned} \tag{2.20}$$

Using these properties, it follows that the expansion coefficients $\tilde{f}_{\ldots}(\mathfrak{R})$ vanish if $\mathfrak{R}=(k,\eta)$ for k half an odd integer, and if $\mathfrak{R}=(s,\frac{1}{2})$.

We conclude this paper by pointing out a particular consequence of Eqs. (2.17), (2.18), and (2.19) and comparing it with a well-known consequence of Eqs. (2.9), (2.10), and (2.11). Let us first state the latter. Let us write $\mathcal{H}'$ for the Hilbert space of functions of a single variable, $f(\beta)$, with the norm defined by

$$(f,f)' \equiv \tfrac{1}{2}\int_0^\infty \sinh\beta d\beta\,|f(\beta)|^2 < \infty. \tag{2.21}$$

Then for any given pair of numbers m, n where both are integral or both are half-odd-integral, the functions $d^{(\mathfrak{R})}_{mn}(\beta)$ form an orthogonal basis for $\mathcal{H}'$, as $\mathfrak{R}$ varies over (k,η) and (s,ϵ). (The discrete series terms are not present if m and n are of opposite signs, or if at least one of them equals $\pm\frac{1}{2}$.) In fact, we have

$$\begin{aligned} f(\beta) &= \int d\mathfrak{R}\mu(\mathfrak{R})\tilde{f}(\mathfrak{R})d^{(\mathfrak{R})}_{mn}(\beta), \\ \tilde{f}(\mathfrak{R}) &= \mu(\mathfrak{R})\cdot\tfrac{1}{2}\int_0^\infty d\beta\,\sinh\beta d^{(\mathfrak{R})}_{mn}(\beta)^* f(\beta), \\ (f,f)' &= \int d\mathfrak{R}\,|\tilde{f}(\mathfrak{R})|^2. \end{aligned} \tag{2.22}$$

The functions $d^{(\mathfrak{R})}_{mn}(\beta)$ are a complete orthogonal system of eigenfunctions of the second-order differential operator defined on $\mathcal{H}'$,

$$Q(m,n;\beta) \equiv -\frac{d^2}{d\beta^2} - \coth\beta\frac{d}{d\beta} + \frac{m^2+n^2-2mn\cosh\beta}{\sinh^2\beta}, \tag{2.23}$$

and they obey

$$\begin{aligned} Q(m,n;\beta)d^{(\mathfrak{R})}_{mn}(\beta) &= q(\mathfrak{R})d^{(\mathfrak{R})}_{mn}(\beta), \\ q(\mathfrak{R}) &= k(1-k) \quad \text{if } \mathfrak{R}=(k,\eta), \\ &= \tfrac{1}{4}+s^2 \quad \text{if } \mathfrak{R}=(s,\epsilon). \end{aligned} \tag{2.24}$$

These results follow from a specialization of Eqs. (2.9), (2.10), and (2.11) to functions $f(g)$ with a simple dependence on α and γ. The analog in the noncompact basis is to consider the special case of column vectors of the type (2.14) in which the dependences of all the entries on ζ and ζ' are taken to be just "plane waves." So we define a Hilbert space $\mathcal{H}''$ to consist of column vectors of the form

$$f = \begin{pmatrix} f(\mu) \\ f_n(\nu) \end{pmatrix}, \quad -2\pi \le \mu \le 2\pi, \quad -\infty < \nu < \infty, \quad n = 0,1,2,3, \tag{2.25}$$

for which

$$\begin{aligned} (f,f)'' &\equiv \frac{1}{8}\int_{-2\pi}^{2\pi} d\mu\,|\sin\mu|\,|f(\mu)|^2 \\ &\quad + \frac{1}{8}\sum_{n=0}^{3}\int_{-\infty}^{\infty} d\nu\,|\sinh\nu|\,|f_n(\nu)|^2 < \infty. \end{aligned} \tag{2.26}$$

Then, given any two real numbers p', p, we obtain a complete orthogonal basis in $\mathcal{H}''$ by considering the column vectors

$$\Psi^{(\mathfrak{R})}_{ba} = \begin{pmatrix} \mathscr{G}^{(\mathfrak{R})}_{ba}(p',p;\mu) \\ \mathfrak{F}^{(\mathfrak{R})}_{ba}(p',p;\nu;n) \end{pmatrix}. \tag{2.27}$$

Here, $\mathfrak{R}$ must now vary over the set (k,η), $\eta=\pm$ and also (s,ϵ), $\epsilon=0,\frac{1}{2}$; and in the latter case, b and a independently assume the values $\pm$ in turn. [The dependence of $\Psi^{(\mathfrak{R})}_{ba}$ on the fixed numbers p', p is not indicated explicitly.] The analog to (2.22) for the space $\mathcal{H}''$ is

$$\begin{aligned} \begin{pmatrix} f(\mu) \\ f_n(\nu) \end{pmatrix} &= \int d\mathfrak{R}\mu(\mathfrak{R})\sum_{ab}\hat{f}_{ba}(\mathfrak{R})\begin{pmatrix} \mathscr{G}^{(\mathfrak{R})}_{ba}(p',p;\mu) \\ \mathfrak{F}^{(\mathfrak{R})}_{ba}(p',p;\nu;n) \end{pmatrix}, \\ \hat{f}_{ba}(\mathfrak{R}) &= \frac{\mu(\mathfrak{R})}{8}\Big(\int_{-2\pi}^{2\pi} d\mu\,|\sin\mu|\,(\mathscr{G}^{(\mathfrak{R})}_{ba}(p'p;\mu))^* f(\mu) \\ &\quad + \sum_{n=0}^{3}\int_{-\infty}^{\infty} d\nu\,|\sinh\nu|\,(\mathfrak{F}^{(\mathfrak{R})}_{ba}(p',p;\nu;n))^* f_n(\nu)\Big) \\ &= \mu(\mathfrak{R})(\Psi^{(\mathfrak{R})}_{ba}, f)'', \\ (f,f)'' &= \int d\mathfrak{R}\sum_{ab}|\hat{f}_{ba}(\mathfrak{R})|^2. \end{aligned} \tag{2.28}$$

The orthogonality and completeness properties can be explicitly stated as[7]

$$(\Psi^{(\mathfrak{R}')}_{b'a'},\ \Psi^{(\mathfrak{R})}_{ba})'' = \delta(\mathfrak{R}',\mathfrak{R})\delta_{b'b}\delta_{a'a}/\mu(\mathfrak{R})\mu(\mathfrak{R}'), \tag{2.29a}$$

$$\begin{aligned} &\frac{1}{8}\int d\mathfrak{R}\,\mu^2(\mathfrak{R})\sum_{ab}\begin{pmatrix} \mathscr{G}^{(\mathfrak{R})}_{ba}(p',p;\mu) \\ \mathfrak{F}^{(\mathfrak{R})}_{ba}(p',p;\nu;n) \end{pmatrix} \\ &\quad \Big(\mathscr{G}^{(\mathfrak{R})}_{ba}(p',p;\mu')^*\ \mathfrak{F}^{(\mathfrak{R})}_{ba}(p',p;\nu';n')^*\Big) \\ &\qquad = \begin{pmatrix} \delta(\mu-\mu')/|\sin\mu|; & 0 \\ 0; & \delta_{nn'}\delta(\nu-\nu')/|\sinh\nu| \end{pmatrix}. \end{aligned} \tag{2.29b}$$

What appears on the right-hand side here is a 5×5 matrix. The increase in complexity of these equations as compared with the corresponding ones in the compact basis is caused by the fact that there is no uniform parametrization of $SU(1,1)$ in the noncompact basis, but that one has to choose different coordinates in each of the five regions. One immediate difference between the spaces $\mathcal{H}'$ and $\mathcal{H}''$ is obvious: for certain choices of m and n, the corresponding basis for $\mathcal{H}'$ may not involve the discrete class UIR's at all, and depending on whether m and n are integers or half-odd-integers, only the integral or half-integral UIR's of $SU(1,1)$ are relevant. In the case of $\mathcal{H}''$, on the other hand, all the UIR's of $SU(1,1)$ that appear in Bargmann's theorem are needed in forming a basis, whatever be the values of the fixed quantities p, p'. The vectors $\Psi^{(\mathfrak{R})}_{ba}$ in $\mathcal{H}''$ also form a complete orthogonal set of eigenfunctions for a certain second-order differential operator that can be defined using the results of I and II. Using the form (2.25) for $\mathcal{H}''$ and the operator $Q(m,n;\beta)$ defined in (2.23), we set up the five-dimensional diagonal matrix operator acting on $\mathcal{H}''$:

$$\hat{Q}(p',p) \equiv \begin{pmatrix} Q(ip',ip;i\mu) & \cdot & \cdot & \cdot & \cdot \\ \cdot & Q(ip',ip;\nu) & \cdot & \cdot & \cdot \\ \cdot & \cdot & Q(ip',-ip;\nu) & \cdot & \cdot \\ \cdot & \cdot & \cdot & Q(ip',ip;\nu) & \cdot \\ \cdot & \cdot & \cdot & \cdot & Q(ip',-ip;\nu) \end{pmatrix}. \tag{2.30}$$

Then we have the result

$$\hat{Q}(p',p)\Psi^{(\mathfrak{R})}_{ba} = q(\mathfrak{R})\Psi^{(\mathfrak{R})}_{ba}. \tag{2.31}$$

Each eigenvalue $k(1-k)$ for $k=1,\frac{3}{2},\ldots$ appears twice for the operator $\hat{Q}(p',p)$, corresponding to $\mathfrak{R}=(k,\pm)$; and each eigenvalue $\frac{1}{4}+s^2$ appears eight times corre-

sponding to $\mathfrak{R}=(s,0)$, $(s,\frac{1}{2})$, $b=\pm$, $a=\pm$. All these results put together clarify the completeness and orthogonality properties of the functions $\mathscr{G}$ and $\mathfrak{F}$ computed in I and II and extended here to the entire group $SU(1,1)$.[8]

*Present address: Centre for Theoretical Studies, Indian Institute of Science, Bangalore 560012, India.

[1]N. Mukunda J. Math. Phys. **10,** 2086 (1969); J. Math. Phys. **10,** 2092 (1969). We refer to these two papers as I, II, respectively. References to the literature on the representation theory of $O(2,1)$ can be found in them. For the representations in a noncompact basis, see also A. O. Barut and E. C. Phillips, Commun. Math. Phys. **8,** 52 (1968); G. Lindblad and B. Nagel, "Continuous bases for unitary irreducible representations of $SU(1,1)$," Department of Theoretical Physics, Royal Institute of Technology, Stockholm, preprint, April 1969.

[2]V. Bargmann, Ann. Math. **48,** 568 (1947).

[3]For other forms of this completeness relation, see G. Lindblad, "Eigenfunction expansions associated with unitary irreducible representations of $SU(1,1)$," Department of Theoretical Physics, Royal Institute of Technology, Stockholm, preprint, February 1970.

[4]N. Mukunda, J. Math. Phys. **8,** 2210 (1967).

[5]It is clear that what we are doing here is an enumeration of the double cosets of $SU(1,1)$ relative to $O(1,1)$.

[6]If, as in Ref. 2, one chooses the ranges $-2\pi \leq \alpha, \gamma \leq 2\pi$ for the sake of symmetry, the group manifold gets covered twice.

[7]The scalar product in $\mathscr{H}''$ follows from (2.26) by polarization.

[8]A direct determination of the complete orthogonal set of eigenfunctions of $\hat{Q}(p', p)$ based on properties of second-order differential operators, along the lines of Ref. 3, would lead to considerable complications.

Motion in radial magnetic fields

André Hautot

University of Liège, Institute of Physics, Sart Tilman, 4000 Liège 1, Belgium
(Received 14 January 1972)

We study the problem of the motion of a charged particle in radial magnetic induction fields of the type $\mathbf{B} = h(\theta)\mathbf{r}/r^3$. A Coulomb electric field is also added. The mechanics of Newton, Einstein, Schrödinger, and Dirac are successively considered in view of exact solutions. The analogies between the four treatments are emphasized.

One way for studying mathematically the fundamental equations of mechanics is to put definite problems in view of *exact* solutions. When one considers the most complicated forces acting on a charged particle which allow an exact integration it is possible to realize exactly the limitations imposed by the actual advancement of analysis. Naturally it is another task to discover whether or not these forces (sometimes rather fantastic) are of interest for physicists. We have neglected this aspect here considering only in this first approach the performing of exact calculations which seem to stay at the limit of what is presently mathematically possible.

Of course such a problem can never be solved in all its generality. It is always necessary to treat separately well distinct classes of potentials. A few exact resolutions in classical mechanics are known[1,2]; in quantum mechanics the problem has been studied by various authors,[3–5] each bringing its own contribution after exposing the presumed complete bibliography. The present paper deals with the search of the general characteristics of the movement of a charged particle in magnetic induction fields of the type

$$\mathbf{B} = h(\theta)\mathbf{r}/r^3 \tag{1}$$

written in spherical coordinates r, θ, φ.

The problem is considered successively in the four fundamental mechanics: Newton's, Einstein's, Schrödinger's, and Dirac's. In view of increasing the generality (and the complexity!) of the calculations we consider that the particle simultaneously experiences an attractive Coulomb electric field. The analogies between the four treatments will be emphasized: it will be shown that they are of a rather strange character.

Notations: We shall denote ϵ and μ_0 the charge and the rest mass of the particle; the relativistic mass will be $\mu = \beta\mu_0$, where $\beta = (1 - v^2/c^2)^{-1/2}$. The Coulomb electric potential is $-(\mu_0/\epsilon)(H/r)$. $\mathbf{v}$ is the velocity of the particle and $\boldsymbol{\gamma}$ its acceleration, i is the symbol of complex numbers.

I. NEWTON'S MECHANICS

Newton's equation is written as

$$\boldsymbol{\gamma} = (\epsilon/\mu_0)\mathbf{v} \wedge h(\theta)\mathbf{r}/r^3 - H\mathbf{r}/r^3. \tag{2}$$

A. The radial integration

The conservation of energy implies

$$v^2 - 2H/r = a \quad (a = \text{const, negative for bound states}).$$

By scalar multiplication of (2) by $\mathbf{r}$ we get $\mathbf{r}\cdot\boldsymbol{\gamma} = -H/r$ from which we deduce $d(\mathbf{r}\cdot\mathbf{v})/dt = a + H/r$.

Remembering that $\mathbf{r}\cdot\mathbf{v} = 1/2\ dr^2/dt$ we have after a classical integration

$$\mathbf{r}\cdot\mathbf{v} = (ar^2 + 2Hr - b)^{1/2} \quad (b = \text{const}).$$

Finally we get

$$\int dt = \int r(ar^2 + 2Hr - b)^{-1/2}\, dr = I(r).$$

The radial motion is therefore described by the exact equation $t - t_0 = I(r) - I(r_0)$. In particular it is independent of the presence of the radial magnetic field. Therefore it coincides with the radial Kepler's movement.

B. The angular integrations

Let us define $\mathbf{P} = \mathbf{r} \wedge \mathbf{v}$ and let us calculate P^2. Denoting α the angle between $\mathbf{r}$ and $\mathbf{v}$ we have

$$P^2 = r^2v^2 \sin^2\alpha$$

and

$$(\mathbf{r}\cdot\mathbf{v})^2 = r^2v^2\cos^2\alpha = ar^2 + 2Hr - b.$$

Adding these two results and taking into account the conservation of energy we arrive at $P^2 = b(>0)$. Thus the constant $\mu_0^2\, b$ is the total angular momentum squared. On another side one has in spherical coordinates:

$$P^2 = r^4(\dot\theta^2 + \sin^2\theta\dot\varphi^2) \tag{3}$$

The overdot denotes time differentiation.

Up to now the calculations are correct for every magnetic field of the species $\mathbf{B} = L(x, y, z)\mathbf{r}$, where the function L is arbitrary. Now we shall confine ourselves to the special form (1).

The field (1) derives from the following potential: $\mathbf{B} = \text{curl}\ \mathbf{A}$ with

$$\mathbf{A} = [g(\theta)/r^2 \sin\theta](-y, x, 0). \tag{4}$$

The functions h and g are connected by: $h(\theta) = g(\theta)\cot\theta + g'(\theta)$. Let us put

$$f(\theta) = (\epsilon/\mu_0)\int h(\theta)\sin\theta d\theta = (\epsilon/\mu_0)\, g(\theta)\sin\theta. \tag{5}$$

After vector multiplication of (2) by $\mathbf{r}$ one computes

$$dP_z/dt = (\epsilon/\mu_0)[h(\theta)/r][\mathbf{v}_z - (\mathbf{r}\cdot\mathbf{v})z/r^2].$$

Remembering that $z = r\cos\theta$ and $v_z = \dot r\cos\theta - r\sin\theta\dot\theta$ we deduce

$$P_z = d - (\epsilon/\mu_0)\int h(\theta)\sin\theta d\theta \quad (d = \text{const}).$$

But $P_z = r^2\sin^2\theta\dot\varphi$ so that we have

$$r^2\sin^2\theta\dot\varphi = d - f(\theta). \tag{6}$$

Equations (3) and (6) allow us to find the two last integrations needed for the complete solution:

$$I=\int \frac{\sin\theta d\theta}{\{P^2\sin^2\theta-[d-f(\theta)]^2\}^{1/2}} = \int \frac{1}{r}(ar^2+2Hr-b)^{-1/2}dr, \tag{7}$$

$$J=\int \frac{[d-f(\theta)]d\theta}{\sin\theta\{P^2\sin^2\theta-[d-f(\theta)]^2\}^{1/2}}=\int d\varphi. \tag{8}$$

The problem is now completely solved. An exact solution exists provided the integrals present in (7) and (8) are elementary. Naturally so long as we do not precise our sights the expression exact solution remains ambiguous. We shall therefore restrict ourselves to a complete solution in term of elementary functions or at most elliptic functions. The θ integrations present in (7) and (8) are elementary in only one case and elliptic in three independent cases. All other cases seem unsolvable or reduce to linear combinations of the former.

1. Elementary integrations

Except in the case $g=\mathcal{K}/\sin\theta$, where **B** vanishes, there is only one possibility:

$$g=\mathcal{K}\cot\theta \rightarrow \mathbf{B}=\mathcal{K}\mathbf{r}/r^3 \text{ (Coulomb magnetic field).}$$

Equations (7) and (8) integrate into elementary functions. For example r and θ are connected by the relation

$$[P^2+(\epsilon\mathcal{K}/\mu_0)^2]^{-1/2}\arcsin\frac{d(\epsilon\mathcal{K}/\mu_0)-[P^2+(\epsilon\mathcal{K}/\mu_0)^2]\cos\theta}{P[P^2+(\epsilon\mathcal{K}/\mu_0)^2-d^2]^{1/2}}$$

$$=P^{-1}\arcsin\frac{Hr-P^2}{r(H^2+aP^2)^{1/2}}.$$

2. Elliptic integrations

It is known from the theory of elliptic functions[6] that the integral $\int R(z,\mathcal{P}^{1/2})dz$, where the function R is rational in z and in $\mathcal{P}^{1/2}=(Az^4+Bz^3+Cz^2+Dz+E)^{1/2}$ is of elliptic kind. Therefore it is sufficient for our purpose to introduce into (7) and (8) a lot of trigonometrical functions in place of $f(\theta)$ and to retain only those which lead to elliptic integrals when these are made algebraic after a suitable change of variables.

Equations (7) and (8) are automatically rational when one puts $y=\tan\theta/2$; however $\mathcal{P}$ is at most of fourth order if $g=\mathcal{K}$ (=const). Likewise if one puts $u=\cos\theta$ we must restrict ourselves to $g=\mathcal{K}\sin\theta$ or $g=\mathcal{K}\tan\theta$. We have never found other possibilities. Let us now review briefly the three cases:

1st case: $g=\mathcal{K} \rightarrow \mathbf{B}=\mathcal{K}\cot\theta\,\mathbf{r}/r^3$.

One puts $y=\tan\theta/2$. Equations (7) and (8) become

$$I=4\int\frac{y\,dy}{(1+y^2)\mathcal{P}^{1/2}},\quad J=\int\frac{d(1+y^2)-2(\epsilon\mathcal{K}/\mu_0)y}{y\mathcal{P}^{1/2}}dy.$$

$\mathcal{P}$ is defined in accordance with the following values:

$$A=E=-d^2,\qquad B=D=4d(\epsilon\mathcal{K}/\mu_0),$$
$$C=4P^2-4(\epsilon\mathcal{K}/\mu_0)^2-2d^2. \tag{9}$$

2nd case: $g=\mathcal{K}\sin\theta \rightarrow \mathbf{B}=2\,\mathcal{K}\cos\theta\,\mathbf{r}/r^3$.

One puts $u=\cos\theta$:

$$I=-\int\frac{du}{\mathcal{P}^{1/2}},\quad J=\int\frac{(\epsilon\mathcal{K}/\mu_0)(1-u^2)-d}{(1-u^2)\mathcal{P}^{1/2}}du,$$

$$A=-(\epsilon\mathcal{K}/\mu_0)^2,\quad B=D=0,$$
$$C=-P^2+2(\epsilon\mathcal{K}/\mu_0)^2-2d(\epsilon\mathcal{K}/\mu_0), \tag{10}$$
$$E=P^2-[d-(\epsilon\mathcal{K}/\mu_0)]^2.$$

3rd case: $g=\mathcal{K}\tan\theta \rightarrow \mathbf{B}=\mathcal{K}(2+\tan^2\theta)\mathbf{r}/r^3$.

One puts $u=\cos\theta$:

$$I=-\int\frac{u\,du}{\mathcal{P}^{1/2}},\quad J=\int\frac{(\epsilon\mathcal{K}/\mu_0)(1-u^2)-du}{(1-u^2)\mathcal{P}^{1/2}}du,$$

$$A=-P^2-(\epsilon\mathcal{K}/\mu_0)^2,\quad B=-D=-2d(\epsilon\mathcal{K}/\mu_0), \tag{11}$$
$$C=P^2-d^2+2(\epsilon\mathcal{K}/\mu_0)^2,\quad E=-(\epsilon\mathcal{K}/\mu_0)^2.$$

To save place we shall omit the complete writing of the solutions in terms of elliptic functions.

Remark: It must be pointed out that conformably to the theory of elliptic functions the elliptic integrals I and J degenerate in elementary integrals if $\mathcal{P}$ has a double root. That is of course possible only if P^2, d, and $\epsilon\mathcal{K}/\mu_0$ (the characteristic parameters of our problem) are suitably connected. For example, in the second case the following condition is needed:

$$P^2=[(\epsilon\mathcal{K}/\mu_0)-d]^2.$$

II. EINSTEIN'S MECHANICS

Einstein's equation is written as:

$$\beta\boldsymbol{\gamma}+(\beta^3/c^2)\mathbf{v}\cdot\boldsymbol{\gamma}\mathbf{v}=(\epsilon/\mu_0)\mathbf{v}\wedge h(\theta)\mathbf{r}/r^3-H\mathbf{r}/r^3. \tag{12}$$

If there was no electric field this equation would be the same as Newton's.[2]

A. The radial integration

The conservation of energy implies

$$\beta-H/(c^2r)=a' \quad (a'=\text{const}). \tag{13}$$

By scalar multiplication of (12) by $\mathbf{r}$ we get

$$\beta\mathbf{r}\cdot\boldsymbol{\gamma}+(\beta^3/c^2)\mathbf{v}\cdot\boldsymbol{\gamma}\mathbf{r}\cdot\mathbf{v}+H/r=0.$$

Hence

$$d(\beta\mathbf{r}\cdot\mathbf{v})/dt=\beta\mathbf{v}^2-H/r.$$

Eliminating β in this equation with the aid of (13) we get

$$\frac{d}{dt}\left((a'c^2r+H)\frac{dr}{dt}\right)=c^4\frac{(a'^2-1)c^2r+a'H}{a'c^2r+H}.$$

The integration of this equation is classical. Putting $w=a'c^2r+H$, one finally arrives at

$$c^2\int dt=\int w[c^2(a'^2-1)w^2+2c^2Hw-b'a'^2]^{-1/2}dw$$
$$(b'=\text{const}),$$

which defines the radial motion by means of elementary functions without reference to the magnetic field: We refind the radial relativistic Kepler's motion.

B. The angular integration

Let us define $\mathbf{J}=\beta\mathbf{r}\wedge\mathbf{v}$ and let us calculate J^2. Working as in Sec. IB one finds

$$J^2=b'/c^4-H^2/a'^2c^2=\text{const}.$$

From another side

$$d\mathbf{J}/dt=(\epsilon/\mu_0)[h(\theta)/r^3]\mathbf{r}\wedge(\mathbf{v}\wedge\mathbf{r})$$

from which we deduce $J_z = d' - f(\theta)$ ($d' = \text{const}$) as in Sec. I B. Summarizing these two results we arrive on account of (3) to the following integrations which allow us to determinate the angular motion:

$$J^2 = \beta^2 r^4(\dot\theta^2 + \sin^2\theta\dot\varphi^2),$$

$$d' - f(\theta) = \beta r^2 \sin^2\theta\dot\varphi.$$

The angular integrals are

$$\int \frac{\sin\theta\, d\theta}{\{J^2\sin^2\theta - [d' - f(\theta)]^2\}^{1/2}} = a'c^2$$

$$\int \frac{dw}{(w-H)[c^2(a'^2-1)w^2 + 2c^2Hw - ba'^2]^{1/2}},$$

$$\int \frac{[d' - f(\theta)]\, d\theta}{\sin\theta\{J^2\sin^2\theta - [d' - f(\theta)]^2\}^{1/2}} = \int d\varphi.$$

The θ and the φ dependence of these equations are exactly identical with those found in Newton's theory so that the same conclusions remain valuable: one field leads to elementary solutions and three others lead to elliptic integrals. These fields have been presented in Sec. I.

III. SCHRÖDINGER'S MECHANICS

A. Separation of the variables

Schrödinger's equation is written as

$$\Delta\psi + 2(\epsilon/\hbar)i\mathbf{A}\cdot\vec{\text{grad}}\psi - (\epsilon/\hbar)^2A^2\psi + (2\mu_0/\hbar^2)(E - V)\psi = 0. \tag{14}$$

Using spherical coordinates one establishes with the aid of (4) that

$$\mathbf{A}\cdot\vec{\text{grad}} = [g/(r^2\sin\theta)]\partial/\partial\varphi,$$

$$A^2 = (g/r)^2.$$

Variables separate in Eq. (14): $\psi = \exp(im\varphi)\textcircled{u}R$. The φ equation immediately integrates into classical imaginary exponential form whilst r and θ equations are (m is the integer magnetic quantum number—we shall only consider positive values of m; calculations are analogous when m is negative):

$$r^2\frac{d^2R}{dr^2} + 2r\frac{dR}{dr} + \frac{2\mu_0}{\hbar^2}r^2\left(E + \frac{\mu_0 H}{r}\right)R + sR = 0, \tag{15a}$$

$$\frac{d^2\textcircled{u}}{d\theta^2} + \cot\theta\frac{d\textcircled{u}}{d\theta} - \left(\frac{m}{\sin\theta} + \epsilon\frac{g}{\hbar}\right)^2\textcircled{u} - s\textcircled{u} = 0. \tag{15b}$$

The physically admissible solutions of (15b) only exist for definite values of the constant parameter s. Then Eq. (15a) is analogous to the radial equation in the hydrogen problem (except the values of s). The discrete energy levels are given by

$$E = -(\mu_0^3H^2/2\hbar^2)[n + 1/2 + (1/4 - s)^{1/2}]^{-2}. \tag{16}$$

Therefore there is only one problem: the resolution of (15b) in view of finding the allowed values for s. An exact solution to this problem exists if the θ equation is of a classical type with polynomial solutions (i.e., Hermite, Laguerre, or Jacobi equation). However as we shall see the field of our investigations is then too narrow so that we shall also admit θ equations which after a suitable change of variables belong to the differential equations trilogy we have previously studied.[7]

The reader will understand later our reasons for doing so. In summary we have to introduce various trigonometrical functions in place of g in Eq. (15b) and to see after a suitable change of variables whether or not it is possible to bring the equation into a classical form (Hermite, Laguerre, or Jacobi) or into one of the three nonusual forms[7]:

$$Df'' + (az^2 + bz + c)f' + (d + ez)f = 0, \tag{17}$$

where $D = z$, $z(z-1)$ or $z(1-z)(\alpha - z)$; we shall call these three equations (17a), (17b) and (17c), respectively. In the theory[7] of Eq. (17) the parameters a, b, and c must verify a very simple condition (j, j', and j'' are integers ≥ 0):

For Eq. (17a):

$$c = -j, \tag{18}$$

For Eq. (17b):

$$c = j, \tag{19a}$$

or

$$a + b + c = -j'. \tag{19b}$$

The second relation deduces from the first when z is replaced by $1 - z$ in (17b).

For Eq. (17c):

$$c = -j\alpha, \tag{20a}$$

or

$$a + b + c = -j'(1 - \alpha), \tag{20b}$$

or

$$a\alpha^2 + b\alpha + c = -j''\alpha(\alpha - 1). \tag{20c}$$

The second relation (resp. the third) deduces from the first when z is replaced by $1 - z$ (resp. by $\alpha - \alpha z$) in (17c).

Here are the solutions we have found to the problem.

1. Classical equations

Excepting the case $g = \mathcal{K}/\sin\theta$, where $\mathbf{B}$ vanishes, there is only one possibility:

$$g = \mathcal{K}\cot\theta \rightarrow \mathbf{B} = \mathcal{K}\mathbf{r}/r^3 \quad \text{(Coulomb field)}.$$

This case leads to Jacobi polynomials if $s = (\epsilon\mathcal{K}/\hbar)^2 - l(l+1)$. We pass over the details since this problem is not new.[8]

2. Nonusual equations of the type (17)

If we put $y = \sin\theta$ in (15b) it may be seen that $g = \mathcal{K}$ ($= \text{const}$) provides a solution to our problem. If we put $u = \cos\theta$ in (15b) it may be seen that $g = \mathcal{K}\sin\theta$ and $g = \mathcal{K}\tan\theta$ are also convenient. We have not succeeded in finding another function g independent of those just mentioned. Let us now review the three cases in greater detail:

1st case: $g = \mathcal{K} \rightarrow \mathbf{B} = \mathcal{K}\cot\theta\mathbf{r}/r^3$.

We put $y = \sin\theta$ in (15b)

$$(1 - y^2)\frac{d^2\textcircled{u}}{dy^2} + \frac{1 - 2y^2}{y}\frac{d\textcircled{u}}{dy} - \left[\frac{m^2}{y^2} + s + \left(\frac{\epsilon\mathcal{K}}{\hbar}\right)^2 + 2m\frac{\epsilon\mathcal{K}}{\hbar}\frac{1}{y}\right]\textcircled{u} = 0. \tag{21}$$

If $\textcircled{u} = y^{-m}T$ we find

$$y(1-y)(-1-y)T'' + [(2-2m)y^2 + (2m-1)]T' + \{2m\epsilon\mathcal{H}/\hbar + [s + (\epsilon\mathcal{H}/\hbar)^2 + m^2 - m]y\}T = 0. \quad (22)$$

Equation (22) is of the type (17c). Only condition (20a) is satisfied with $j = 2m-1$.

2nd case: $g = \mathcal{H}\sin\theta \rightarrow \mathbf{B} = 2\mathcal{H}\cos\theta\mathbf{r}/r^3$.

We put $u = \cos\theta$ in (15b):

$$(1-u^2)\frac{d^2\textcircled{u}}{du^2} - 2u\frac{d\textcircled{u}}{du} - \left[\frac{m^2}{1-u^2} + \left(\frac{\epsilon\mathcal{H}}{\hbar}\right)^2(1-u^2) + s + \frac{2m\epsilon\mathcal{H}}{\hbar}\right]\textcircled{u} = 0. \quad (23)$$

If $\textcircled{u} = (1-u^2)^{-m/2}\exp[-(\epsilon\mathcal{H}/\hbar)u]T$ and $u = 2v-1$ for the sake of convenience:

$$v(v-1)T'' + [-4(\epsilon\mathcal{H}/\hbar)v^2 + (4\epsilon\mathcal{H}/\hbar - 2m + 2)v + (m-1)]T' + [(2\epsilon\mathcal{H}/\hbar + s + m^2 - m) + 4(\epsilon\mathcal{H}/\hbar)(m-1)v]T = 0. \quad (24)$$

Equation (24) is of the type (17b). Both conditions (19a) and (19b) are satisfied with $j = j' = m-1$.

3rd case: $g = \mathcal{H}\tan\theta \rightarrow \mathbf{B} = \mathcal{H}(2 + \tan^2\theta)\mathbf{r}/r^3$.

We put $u = \cos\theta$ in (15b):

$$(1-u^2)\frac{d^2\textcircled{u}}{du^2} - 2u\frac{d\textcircled{u}}{du} - \left[\frac{m^2}{1-u^2} + 2m\,\frac{\epsilon\mathcal{H}}{\hbar}\,\frac{1}{u} + \left(\frac{\epsilon\mathcal{H}}{\hbar}\right)^2\frac{1-u^2}{u^2} + s\right]\textcircled{u} = 0. \quad (25)$$

If $\textcircled{u} = (1-u^2)^{-m/2}u^{\sigma}T$ with $\sigma^2 - \sigma - (\epsilon\mathcal{H}/\hbar)^2 = 0$, we have

$$u(1-u)(-1-u)T'' + [2(\sigma+1-m)u^2 - 2\sigma]T' + \{2m\epsilon\mathcal{H}/\hbar + [-m + s + m^2 + 2\sigma(1-m)]u\}T = 0. \quad (26)$$

Equation (26) is of the type (17c). Both conditions (20b) and (20c) are satisfied with $j' = j'' = m-1$.

Digression: In the three cases T obeys an equation like (17). Before going on let us make a digression about the solutions of (17). We recall and extend the results of Ref. 7. Because of the condition of finiteness we shall restrict ourselves to polynomial T functions:

$$T = \sum_0^n \lambda_k z^k.$$

A first polynomial condition is

$$\begin{aligned} e &= -an && \text{for (17a) and (17b),} \\ e &= -n(n+a-1) && \text{for (17c).} \end{aligned} \quad (27)$$

The recurrence equation which gives λ_k is $R_k\lambda_{k-1} + S_k\lambda_k + T_k\lambda_{k+1} = 0$ $(k = 0, \ldots, n)$. The coefficients R_k, S_k and T_k are defined in Ref. 7. The last equation is compatible only if

$$\begin{vmatrix} S_0 & T_0 & & \\ R_1 & S_1 & T_1 & \\ \bullet & \bullet & \bullet & \\ \bullet & \bullet & \bullet & \\ \bullet & \bullet & \bullet & \\ & R_{n-1} & S_{n-1} & T_{n-1} \\ & & R_n & S_n \end{vmatrix} = 0. \quad (28)$$

This determinant equation may be fulfilled in two different ways provided a, b, and c are well-connected [see Eqs. (18), (19), and (20)]. Indeed $T_j = 0$ so that two possibilities exist to satisfy (28):

1st possibility:

$$\begin{vmatrix} S_0 & T_0 & & \\ R_1 & S_1 & T_1 & \\ \bullet & \bullet & \bullet & \\ \bullet & \bullet & \bullet & \\ \bullet & \bullet & \bullet & \\ & R_{j-1} & S_{j-1} & T_{j-1} \\ & & R_j & S_j \end{vmatrix} = 0,$$

2nd possibility: $\lambda_0 = \lambda_1 = \ldots = \lambda_j = 0$,

$$\begin{vmatrix} S_{j+1} & T_{j+1} & & \\ R_{j+1} & S_{j+1} & T_{j+1} & \\ \bullet & \bullet & \bullet & \\ \bullet & \bullet & \bullet & \\ \bullet & \bullet & \bullet & \\ & R_{n-1} & S_{n-1} & T_{n-1} \\ & & R_n & S_n \end{vmatrix} = 0 \quad (29)$$

so that

$$T = z^{j+1} \times \text{polynomial of degree } n-j-1. \quad (30)$$

We have seen[7] that with the first possibility the solution of (17) appear as linear combinations of Hermite, Laguerre, or Jacobi polynomials. The second possibility is more difficult to deal with because the order of the determinant (29) increases with the order of the polynomial. Yet only the second possibility is interesting for our purpose because of the divergent from of $\textcircled{u}$ in (21), (23), and (25): indeed it will be shown on each particular case that on account of (30) $\textcircled{u}$ is finally regular with the second possibility whilst it would be divergent with the first.

B. Resolution of the differential equations

We only study the three magnetic induction fields mentioned above. As we shall see no physically admissible solution exist unless the parameters characterizing the problem are suitably connected. In particular the parameter $\mathcal{H}$ (which has the dimensions of a magnetic pole) cannot take arbitrary values.

1st case: $\mathbf{B} = \mathcal{H}\cot\theta\mathbf{r}/r^3$.

We must deal with Eq. (22). Because $j = 2m-1$ in that case we have on account of (30)

$$T = y^{2m}P^{(\nu)},$$

where $P^{(\nu)}$ denotes a polynomial of degree ν. The first polynomial condition [see (27)] gives the allowed values for s:

$$s = -(\epsilon\mathcal{H}/\hbar)^2 - (m+\nu)(m+\nu+1). \quad (31)$$

If $\mathcal{H} = 0$ we recover the classical hydrogen values $s = -l(l+1)$ if we set $l = m + \nu$. The second polynomial condition [see the determinant condition (29)] is

$$\begin{vmatrix} S_{2m} & T_{2m} & & \\ R_{2m+1} & S_{2m+1} & T_{2m+1} & \\ \bullet & \bullet & \bullet & \\ \bullet & \bullet & \bullet & \\ \bullet & \bullet & \bullet & \\ & R_{2m+\nu-1} & S_{2m+\nu-1} & T_{2m+\nu-1} \\ & & R_{2m+\nu} & S_{2m+\nu} \end{vmatrix} = 0. \tag{32}$$

For each value of ν it gives the allowed values for the parameter $\mathcal{K}$. For example, $\nu=1$ (a simple case): the determinant (32) is of order two and the values of R_k, S_k and T_k are given in Ref. 7; one finds

$$\begin{vmatrix} 2m\epsilon\mathcal{K}/\hbar & -2m-1 \\ -2m-2 & 2m\epsilon\mathcal{K}/\hbar \end{vmatrix} = 0,$$

i.e.,

$$\mathcal{K} = (\hbar/2\epsilon)\{[(2m+1)(2m+2)]^{1/2}/m\}.$$

We deduce the value of s in this special case and we find the discrete energy levels numbered by the integer n [see Eq. (16)]:

$$E = -(\mu_0^3 H^2/2\hbar^2)(n+1/2+\{1/4+[(m+1)/2m^2](2m^3 + 4m^2+2m+1)\}^{1/2})^{-2}.$$

Each value of ν must be analyzed separately in the same manner. Writing the determinant condition (32) in each cases we can at least, when ν is small, list all the values of $\mathcal{K}$ which allow us to solve completely Schrödinger's equation. Inversely the value of $\mathcal{K}$ being fixed (among the allowed values of course) the corresponding values of m and ν deduce like that of s and finally (16) allows us to construct the energy levels spectrum (with $n=0,1,2,\ldots$). Since m and s are related to the angular momentum proper values we conclude by saying that a given allowed value of $\mathcal{K}$ automatically forbids arbitrary angular momentum states.

2nd case: $\mathbf{B} = 2\mathcal{K}\cos\theta\mathbf{r}/r^3$.

We must deal with Eq. (24). In fact this case is very different from the two others because it does not lead to a discrete energy levels spectrum. Indeed the first polynomial condition [see (27)] can never be satisfied; one must have on account of (30)

$$T = v^m(v-1)^m P^{(\nu)}.$$

So that T would be a polynomial of degree $\nu+2m$; therefore, (27) would need $4(\epsilon\mathcal{K}/\hbar)(m-1) = 4(\epsilon\mathcal{K}/\hbar)(\nu+2m)$, which is impossible. We shall try to see in the final discussion the reasons for that special demeanor.

3rd case: $\mathbf{B} = \mathcal{K}(2+\tan^2\theta)\mathbf{r}/r^3$.

We must deal with Eq. (26). Because $j' = j'' = m-1$ in that case we have on account of (30)

$$T = (1-u^2)^m P^{(\nu)}. \tag{33}$$

The first polynomial condition [see (27)] gives the allowed values for s:

$$s = -(\nu+2\sigma+m)(\nu+m+1).$$

If $\mathcal{K}=0$ we recover the classical value $s=-l(l+1)$ by setting $l=\nu+m$. To obtain the polynomial determinant condition we must introduce (33) into (26). Just like in the first case each value of ν must be analyzed separately. The determinant condition gives in each cases the allowed $\mathcal{K}$ values. Let us treat again the special case $\nu=1$; the determinant is of order two:

$$\begin{vmatrix} 2m\epsilon\mathcal{K}/\hbar & -2\sigma \\ -2m-2\sigma-2 & 2m\epsilon\mathcal{K}/\hbar \end{vmatrix} = 0.$$

Remembering the definition $\sigma[\sigma^2-\sigma-(\epsilon\mathcal{K}/\hbar)^2=0]$, one finds the allowed $\mathcal{K}$ values:

$$\mathcal{K} = (\hbar/\epsilon)\{[(m+2)(m^2+m+1)]^{1/2}/(m^2-1)\}.$$

The corresponding s values follow and so do the energy levels

$$E = -(\mu_0^3 H^2/2\hbar^2)(n+1/2+\{1/4+[(m+2)/(m^2-1)](m^3 + 3m^2+m+1)\}^{1/2})^{-2}.$$

Remark: Before concluding this section it is interesting to recall what has been done. Schrödinger's equation seems completely solvable in only one case. In three other cases it is conditionally solvable: $\epsilon\mathcal{K}/\hbar$ must be correctly connected with the angular momentum quantum numbers so that $\mathcal{K}$ is irrational in unit $\hbar/\epsilon$.

IV. DIRAC'S MECHANICS

In this section we deal with very complicated equations. Our first care must be the separation of the spherical variables in Dirac's equation. We have shown elsewhere that this operation is performed in the easiest way by using the quaternionic formalism.[7,9] To save place here the reader is referred to Ref. 7, Sec. IB to discover the details of the quaternionic procedure: it is shown that the φ equation simply leads to trigonometrical functions and that the radial equation is analogous to that of the hydrogen atom problem except that $(J+1/2)^2$ is replaced by the parameter $-\lambda^2$ to be determined. So we deduce that the energy levels are given by the well-known hydrogen formula

$$E = \mu_0 c^2\{1+(\mu_0 H/c\hbar)^2\{n+[-\lambda^2-(\mu_0 H/c\hbar)^2]^{1/2}\}^{-2}\}^{-1/2}. \tag{34}$$

The θ equation alone depends upon the presence of the magnetic potential. Therefore we have only to solve the following coupled system:

$$\frac{dT_1}{d\theta} = -\lambda T_4 - \frac{(2m+1)+\cos\theta}{2\sin\theta}T_1 - (\epsilon g/\hbar)T_1,$$

$$\frac{dT_4}{d\theta} = -\lambda T_1 + \frac{(2m+1)-\cos\theta}{2\sin\theta}T_4 + (\epsilon g/\hbar)T_4. \tag{35}$$

This system is exactly solvable by means of elementary functions in the case $g=\mathcal{K}\cot\theta$, i.e., for the field of a magnetic pole. We pass over the details since the problem has been investigated by Harish-Chandra.[8] It is conditionally solvable in three cases:

1st case: $g=\mathcal{K} \rightarrow \mathbf{B} = \mathcal{K}\cot\theta\mathbf{r}/r^3$.

We make the substitution $z=\exp(i\theta)$ in (35) and we decouple the system to obtain two second-order rational equations. We shall only deal with the T_1 equation since when T_1 is known, T_4 deduces by (35); it is written with arbitrary g function ($g'=dg/d\theta$):

$$z^2\frac{d^2T_1}{dz^2} + \frac{2z^3}{z^2-1}\frac{dT_1}{dz} + \left[(\lambda^2+1/2) - \frac{4m^2z^2}{(z^2-1)^2} - \frac{2mz}{(z-1)^2}\right.$$

$$-\left(\frac{1}{4}\right)\left(\frac{z+1}{z-1}\right)^2+2i(2m+1)\ \frac{\epsilon}{\hbar}\ g\frac{z}{z^2-1}+\left(\frac{\epsilon g}{\hbar}\right)^2 - \frac{\epsilon g'}{\hbar}\Bigg]T_1=0.$$

This equation is too general, since we only deal with $g=\mathcal{K}$ (=const) so that $g'=0$. Let us set

$$T_1=(z-1)^{-m-1}(z+1)^{-m}z^{\sigma}V_1,$$

where $\sigma^2-\sigma+\lambda^2+(\epsilon\mathcal{K}/\hbar)^2+1/4=0$. We find

$$z(1-z)(-1-z)V_1''+[(2\sigma-4m)z^2-2z-2\sigma]V_1' +\{[(2m+1)(1+2i\epsilon\mathcal{K}/\hbar)-2\sigma] +(4m^2+2m-4m\sigma)z\}V_1=0. \quad (36)$$

Equation (36) is of the type (17c). Both conditions (20b) and (20c) are satisfied with $j'=2m+1$ and $j''=2m-1$. Therefore, we have on account of (30)

$$V_1=(z-1)^{2m+2}(z+1)^{2m}P^{(\nu)},$$

where $P^{(\nu)}$ is a polynomial of degree ν. The first polynomial condition [see (27)] gives the allowed values for λ:

$$4m^2+2m-4m\sigma=-(\nu+4m+2)(\nu+2\sigma+1), \quad (37)$$

which reduces after simplifications to

$$2\sigma+2m+\nu+1=0. \quad (38)$$

If $\mathcal{K}=0$ we recover the classical hydrogen value $\lambda^2=-(J+1/2)^2$ if we set $J=1/2(\nu+2m+1)$ with ν even. When ν is odd no classical equivalent exists. To obtain the polynomial determinant condition we must introduce (37) in (36). Each value of ν must be analyzed separately. Let us treat the case $\nu=1$; the determinant is of order $\nu+1=2$:

$$\begin{vmatrix} 2i(2m+1)(\epsilon\mathcal{K}/\hbar)-1 & 2m+2 \\ -2m-2 & 2i(2m+1)(\epsilon\mathcal{K}/\hbar)+1 \end{vmatrix}=0,$$

i.e.,

$$\mathcal{K}=\frac{\hbar}{2\epsilon}\left(\frac{2m+3}{2m+1}\right)^{1/2}.$$

The fact that these values are not identical with those found in the nonrelativistic theory may not surprise since in Dirac's theory the charged particle automatically carries a magnetic moment.

We deduce from (38) and the definition of σ the value of λ^2 in this special case ($\nu=1$):

$$\lambda^2=-[(m+1)^2(2m+3)]/(2m+1).$$

Finally, on account of (34) the energy levels are

$$E=\mu_0c^2[1+(\mu_0H/c\hbar)^2(n+\{[(m+1)^2(2m+3)/(2m+1)] -(\mu_0H/c\hbar)^2\}^{1/2})^{-2}]^{-1/2}.$$

Let us remark in passing that the presence of the magnetic field allows us to consider a strong Coulomb field. In the hydrogen theory one is limited by the condition $\mu_0H/c\hbar\leq 1$ equivalent to $Z<137$ if Z is the atomic number but here the limitation is not so strict: $\mu_0H/c\hbar\leq 3$ equivalent to $Z<411$.

2nd case: $g=\mathcal{K}\sin\theta\rightarrow \mathbf{B}=2\mathcal{K}\cos\theta\mathbf{r}/r^3$.

We make the substitution $u=\cos^2\theta/2$ in (35) and we decouple the system. We again only retain the T_1 equation; for arbitrary g we have as follows:

$$u(1-u)\frac{d^2T_1}{du^2}-(2u-1)\frac{dT_1}{du}-\Bigg[(\lambda^2+1/2)+\frac{(m+u)^2}{4u(1-u)} +\frac{\epsilon g}{\hbar}\ \frac{2m+1}{\sin\theta}+\left(\frac{\epsilon g}{\hbar}\right)^2-\ \frac{\epsilon g'}{\hbar}\Bigg]T_1=0. \quad (39)$$

Here we deal with the special case $g=\mathcal{K}\sin\theta$ so that $g'=\mathcal{K}\cos\theta$. We set

$$T_1=u^{-m/2}(1-u)^{-(m+1)/2}\exp[-2(\epsilon\mathcal{K}/\hbar)u]V_1,$$

where

$$u(u-1)V_1''+\{-4(\epsilon\mathcal{K}/\hbar)u^2+[4(\epsilon\mathcal{K}/\hbar)-2m+1]u+(m-1)\}V_1' +\{(\lambda^2+m^2+4\epsilon\mathcal{K}/\hbar)-4(m-1)(\epsilon\mathcal{K}/\hbar)u\}V_1=0.$$

This equation is of the type (17b). Both conditions (19a) and (19b) are satisfied with $j=m-1$ and $j'=m$. However it is impossible to set here

$$V_1=u^m(1-u)^{m+1}P^{(\nu)},$$

where $P^{(\nu)}$ would be a polynomial of degree ν. Indeed the first polynomial condition [see (27)] can never be satisfied. Therefore no quantization of energy exists. The situation was similar in Schrödinger's theory.

3rd case: $g=\mathcal{K}\tan\theta\rightarrow \mathbf{B}=\mathcal{K}(2+\tan^2\theta)\mathbf{r}/r^3$.

We again make the substitution $u=\cos^2\theta/2$ in (35) so that we can start with (39) where $g'=\mathcal{K}/\cos^2\theta$. Let us put in (39):

$$T_1=(2u-1)^{\epsilon\mathcal{K}/\hbar}u^{-m/2}(1-u)^{-(m+1)/2}V_1.$$

One deduces

$$u(1-u)(1/2-u)V_1''+[(2\epsilon\mathcal{K}/\hbar-2m+1)u^2+(2m-3/2 -2\epsilon\mathcal{K}/\hbar)u-(m-1)/2]V_1'+\{-(\lambda^2+m^2)/2 +[\lambda^2+m^2+2(m+1)\epsilon\mathcal{K}/\hbar]u\}V_1=0. \quad (40)$$

Equation (40) is of the type (17c). Both conditions (20a) and (20b) are satisfied with $j=m-1$ and $j'=m$. Therefore, we have on account of (30)

$$V_1=u^m(1-u)^{m+1}P^{(\nu)}, \quad (41)$$

where $P^{(\nu)}$ is a polynomial of degree ν.

The first polynomial condition [see (27)] gives the allowed values for λ^2:

$$\lambda^2+m^2+2(m+1)\epsilon\mathcal{K}/\hbar=-(\nu+2m+1)(\nu+2\epsilon\mathcal{K}/\hbar+1). \quad (42)$$

If $\mathcal{K}=0$ we recover $\lambda^2=-(J+1/2)^2$ as in the hydrogen theory provided one sets $J=\nu+m+1/2$.

To obtain the polynomial determinant condition we must introduce (41) into (40). Each value of ν must be analyzed separately. Let us treat again the case $\nu=1$; the determinant is of order two:

$$\begin{vmatrix} m\epsilon\mathcal{K}/\hbar+m+3\epsilon\mathcal{K}/\hbar+3/2 & (m+1)/2 \\ -2m-3-2\epsilon\mathcal{K}/\hbar & (m+1)(\epsilon\mathcal{K}/\hbar-1) \end{vmatrix}=0,$$

where account has been taken of (42).

We deduce

$$\mathcal{K}=(\hbar/2\epsilon)(m+3)^{-1}.$$

These are the $\mathcal{K}$-allowed values in the special case $\nu=1$. From (42) we deduce the corresponding values of λ^2 and finally the energy levels by (34):

$$E=\mu_0c^2(1+(\mu_0H/c\hbar)^2\{n+[(m^3+7m^2+19m+15)/(m+3)-(\mu_0H/c\hbar)^2]^{1/2}\}^{-2})^{-1/2}.$$

We remark like in the first case that the presence of the special magnetic field here considered allows to deal with strong Coulomb fields. The limitation is $\mu_0H/c\hbar \leqslant 5$ equivalent to $Z<685$ in the hydrogen theory.

V. DISCUSSION

It is time to compare the solutions obtained in the four fundamental mechanics for the motion of a charged particle in a radial magnetic field of the type (1). The fact that there are analogies between the four treatments is not of course surprising. Let us emphasize them. In nonquantum mechanics: there is only one field (1) which leads to elementary integrations. Three other fields lead to elliptic integrals. All others seem to be more complicated (we shall say unsolvable). In quantum mechanics there is only one field which leads to classical differential equations. Three other fields lead to non-usual differential equations of the type (17). All other fields seem to be unsolvable in that frame. What is very remarkable is that these fields are the same in the four mechanics. That analogy is purely formal and may be pursued in the following way: the final θ integration (resp. the final θ-differential equation) needs a suitable change of variables. In the four treatments this change of variables is analogous when one considers the cases $g_2=\mathcal{K}\sin\theta$ and $g_3=\mathcal{K}\tan\theta$ but is different for $g_1=\mathcal{K}$. Let us now look at more physical analogies.

When one deals with the solutions of the quantum equations in the cases g_1, g_2, and g_3 one may say:

(1) For what regards the cases g_1 and g_3 the problem is conditionally entirely solvable (with discrete energy levels): the parameter $\mathcal{K}$ entering into the definition of g may only take well-chosen values. When this value is fixed the angular momentum quantum numbers m and ν are also fixed at definite values. Such a demeanor is not entirely new in Schrödinger's theory: it is known that definite central electric potentials, like for example $V=V_0\exp(-r/d)$ or $V=V_0\tanh^2(r/d)$, lead to solvable equations if and only if the angular momentum quantum number $l=0$.[10]

(2) The case g_2 is very different since the energy spectrum is not discrete on account of the absence of polynomial solutions.

In nonquantum mechanics the situation is quite analogous:

(1) We have already pointed out in remark (1) that when $\epsilon\mathcal{K}/\mu_0$, P and P_z are suitably connected the elliptic integrals may degenerate into elementary ones. So both integrals I and J [see (7) and (8)] become of the type $\int R[z,(az^2+bz+c)^{1/2}]dz$, where R denotes a rational function. It is known that following the sign of a and c the dependence upon z of that integral may be inverse trigonometrical or logarithmical. It might be shown with the aid of (9), (10), and (11) that in the cases of g_1 and g_3 the solution is not logarithmical so that the motion is a stable orbit (i.e., $r_{\min}<r<r_{\max}$).

(2) Quite the contrary in the case of g_2 the integral is always logarithmical so that the motion is not stable: the particle falls on the center. As Gupta[11] pointed out it does not correspond discrete energy levels to such a spiral orbit in the equivalent quantum problem.

We may conclude be saying that the quadruple treatment of a same problem in four different mechanics exhibits expected physical analogies and also formal analogies which are sometimes of a strange kind. A stable motion is possible in a Coulomb magnetic field but also in the fields $\mathbf{B}=\mathcal{K}\cot\theta\mathbf{r}/r^3$ or $\mathbf{B}=\mathcal{K}(2+\tan^2\theta)\mathbf{r}/r^3$ provided the value of $\mathcal{K}$ is allowed.

[1] E. T. Whittaker, *A Treatise on the Analytic Dynamics of Particles* (Cambridge U. P., Cambridge, 1960).
[2] A. Hautot, Physica **58**, 37 (1972).
[3] N. D. Sengupta, *Satyendranath Bose, 70th Birthday Commemoration Volume* (Calcutta, 1954), p. 55.
[4] L. Lam, Phys. Lett. A **31**, 406 (1970).
[5] A. Hautot, Phys. Lett. A **35**, 129 (1971).
[6] E. T. Whittaker and G. N. Watson, *Modern Analusis* (Cambridge U.P., Cambridge, 1935).
[7] A. Hautot, J. Math. Phys. **13**, 710 (1972).
[8] Harish-Chandra, Phys. Rev. **74**, 883 (1948).
[9] A. Hautot, Physica (Utr.) **48**, 609 (1970).
[10] P. M. Morse and H. Feshback, *Methods of Theoretical Physics* (McGraw-Hill, New York, 1953).
[11] N. D. Sengupta, Indian J. Pure Appl. Phys. **8**, 765 (1970).

Errata: Off-shell T matrix corresponding to a sum of Coulomb and separable potentials by expansion in O(4) harmonics [J. Math. Phys. 12, 1379 (1971)]

W. W. Zachary

Naval Research Laboratory, Washington, D. C. 20375
(Received 26 July 1973)

Ž. Bajzer has kindly informed me of the following errors in the above paper:

(1) In Eq. (10), the second term on the right-hand side should read:

$$-\frac{1}{2p_0}\sum_{\mu'} V^S_{\mu\mu'} T^C_{\mu',\nu}.$$

(2) The decomposition given in Eq. (13) is not "automatic" as claimed, but is merely convenient.

(3) In Eq. (36b), the hypergeometric function in the second term should read:

$${}_2F_1\left(1, -i\mu; 2 - i\mu; \left(\frac{\beta - ik}{\beta + ik}\right)\left(\frac{p - k}{p + k}\right)\right)$$

and the third term should read

$$\frac{2k}{\beta + ik}\,|\,\Gamma(1 + i\mu)\,|^2 e^{\pi\mu}\left(\left(\frac{\beta - ik}{\beta + ik}\right)\left(\frac{p - k}{p + k}\right)\right)^{i\mu}.$$

Errata: Reduction of the Poincaré group with respect to the Lorentz group [J. Math. Phys. 13, 1585 (1972)]

S. W. MacDowell and Ralph Roskies

Department of Physics, Yale University, New Haven, Connecticut 06520
(Received 27 June 1973)

The following misprints should be corrected:

(1) Author's name: S. W. MacDowell and Ralph Roskies.

(2) In Eq. (4.2) replace $(z_2 - z_2)$ by $(z_2 - z_3)$.

(3) In the first line of Eq. (4.11) where it is written $(\lambda_3\lambda_4)$ should be $(\lambda_3 + \lambda_4)$

(4) On the line following Eq. (4.14) replace (4.10) by (4.12).

(5) On the line preceding Eq. (5.3) the factor $\eta^{1-i\lambda j}$ should be $\eta^{1+i\lambda j}$.

(6) In the second line of Eq. (5.3) the factor $(N_j \cdot p_j)$ should be replaced by $(N_j \cdot p_j)^{-1-i\lambda j}$.

(7) In the first line on p. 1590 replace (5.2) by (5.3).

(8) In Appendix C, Eq. (C5) should read:

$$J = \int_0^\infty [(1/m)(p_0 - p) + (1/m)(p_0 + p)\rho^2]^{-1-i\lambda_2}\,\pi d\rho^2$$
$$= (\pi/i\lambda_2)[(1/m)(p_0 + p)]^{-1+i\lambda_2}.$$

(9) In Eq. (C7) replace d^3p/zp_0 by $d^3p/2p_0$.

Erratum: Approximate functional integral methods in statistial mechanics. I. Moment expansions [J. Math. Phys. 13, 1681 (1973)]

Armand Siegel and Terence Burke*

Department of Physics, Boston University, Boston, Massachusetts 02215
(Received 30 May 1973)

The discussion of convergence properties of the expansion on p. 1683 of this article is incorrect. The discussion given would serve to guarantee convergence only if the condition were added that the potential function be bounded from above. While this condition would not prevent the expansion from being useful, it is fortunate that convergence can in fact be proved without it: Kac, in Ref. 4 of our article, has a convergence proof which will serve this purpose. He deals with the term-by-term expectation value of the expansion of the exponential function of the time integral of the (negative of the) *non*-centralized potential, but our centralized potential fits his conditions on the potential—except that he requires that $V \geq 0$, but this can be trivially changed to our condition of boundedness from below; thus we must retain the latter condition. He is able to eliminate the condition of boundedness from above. Finally, it is important to note that Kac proves not only that the expansion converges, but that it converges to the desired Wiener integral.

*Present address: Department of Physics, Texas A. & M. University, College Station, Texas 77843.