JOURNAL OF MATHEMATICAL PHYSICS VOLUME 12, NUMBER 3 MARCH 1971

Unitary Irreducible Representations of $SU(2, 2)$. III. Reduction with Respect to an Iso-Poincaré Subgroup*†

TSU YAO
Department of Physics, University of Pittsburgh, Pittsburgh, Pennsylvania 15213

(Received 3 August 1970)

The unitary irreducible representations of $SU(2, 2)$, the covering group of the conformal group, are reduced with respect to an iso-Poincaré subgroup $E(3, 1)$. Explicit representations of the 15 generators of $SU(2, 2)$ in terms of differential operators in the function space of $|\eta, \xi, \epsilon; s, t, m\rangle$, $|\omega, \xi, \epsilon; r, t, m\rangle$, and $|\xi, \epsilon; t, m\rangle$, the basis vectors for "timelike," "spacelike," and "lightlike" UIR's of $E(3, 1)$, respectively, are given. The matrix elements between a basis vector in the maximal compact subgroup $SU(2) \times SU(2) \times U(1)$ and a basis vector in $E(3, 1)$ are calculated for all 14 different classes of UIR's of $SU(2, 2)$. We find that two classes are reducible with respect to "lightlike" representations, two classes are reducible with respect to "timelike" representations, eight classes are reducible with respect to "spacelike" representations, and two classes contain both "timelike" and "spacelike" representations of $E(3, 1)$.

1. INTRODUCTION

The noncompact group $SU(2, 2)$, which is the covering group of the conformal group, or $O(4, 2)$, has attracted some attention in recent years among particle physicists. There are vigorous attempts at the study of broken conformal and chiral symmetry,[1] hadronic electromagnetic form factors,[2] and in the utilization of the conformal group in quantum electrodynamics calculations.[3] Our interest in this paper is a group theoretic and algebraic one, namely, the study of the unitary irreducible representations of $SU(2, 2)$ and their reduction with respect to a subgroup $E(3, 1)$ which is isomorphic to the Poincaré group.

The more conventional and traditional approach to the reduction of a noncompact group has been the study of its maximal compact subgroup, in which algebraic operations play the dominant role. The determination of the eigenvalues of the Casimir operators for the various unitary irreducible representations is usually most conveniently carried out in this framework. On the other hand, the reduction of a noncompact group with respect to its noncompact subgroups has only recently been given some attention.[4] This study leads to coupled partial differential equations, and their solutions are the various special functions. The two approaches are quite different in their outlook and methodology, but they complement each other, making certain computations much easier or possible in one and not in the other method.

Section 2 starts with the Lie algebra of the group in both its pseudo-unitary $SU(2, 2)$ and pseudo-orthogonal $O(4, 2)$ notations. It then proceeds to the introduction of a noncompact subgroup which is isomorphic to the Poincaré group. In the $O(4, 2)$ language we select three spacelike coordinates 1, 2, 5 and one timelike coordinate 6 for special attention. (We do not use the ordinary space–time coordinates 1, 2, 3, 0, in order to avoid confusion and eliminate prejudice. Of course, it is trivial to change from a 1, 2, 5, 6 space to a 1, 2, 3, 0 space.) We then reduce this iso-Poincaré subgroup $E(3, 1)$ with respect to its noncompact subgroups $E(3)$ and $E(2)$, the Euclidean groups in three and two dimensions.

Section 3 introduces basis vectors $|\eta, \xi, \epsilon; s, t, m\rangle$ of a unitary irreducible representation (UIR from now on) of $E(3, 1)$. Here, we deal with the "timelike" representations, and η can be regarded as the "mass"; ξ, the magnitude of the "3-momentum"; ϵ, the projection of "momentum" in the (1, 2) plane; s, the "spin"; t, the "helicity"; and m is the component of "angular momentum" along the 5 direction. The quotation marks are used to denote all quantities in the 1, 2, 5, 6 space. The generators $\mathfrak{P}_\mu$ and $\mathfrak{L}_{\mu\nu}$ and the covariant spin operators W_μ are represented as simple functions or differential operators in the function space $|\eta, \xi, \epsilon; s, t, m\rangle$.

Section 4 discusses the representation of the remaining five generators of $SU(2, 2)$. The dilatation operator L_{03} is easily represented, but the four conformal operators $\mathfrak{K}_\mu$ are much more complicated. They involve second-order differential operators in η, ξ, and ϵ with $\Delta t = 0, \pm 1$, and $\Delta s = 0, \pm 1$; the representations also involve three yet-to-be-determined functions $\alpha(s)$, $\beta(s)$, and $\gamma(s)$ which are related to the eigenvalues of the three Casimir operators C_2, C_3, and C_4 of $SU(2, 2)$. The explicit representations of $\mathfrak{K}_\mu$ and the determination of $\alpha(s)$, $\beta(s)$, and $\gamma(s)$ are some of the main results of the present article. The algebraic details are, however, relegated to Appendix C.

Section 5 deals with the reduction of $SU(2, 2)$ with respect to $E(3, 1)$ in the "spacelike" region. The results of Sec. 4 can be bodily carried over here by

the substitution of η by $i\omega$ and $s(s+1)$ by $r(r-1)$, where r may be complex. The change of the little group from $SU(2)$ to $SU(1,1)$ makes Bargmann's classification of the UIR's of $SU(1,1)$ particularly relevant.

Section 6 deals with the reduction of $SU(2,2)$ with respect to $E(3,1)$ in the "lightlike" region. The discrete "spin" case is similar to that of Sec. 4, only much simpler. The continuous "spin" case is studied and found to be *not* contributing to the UIR's of $SU(2,2)$.

Section 7 gives an exhaustive study of the matrix elements between a basis vector in the maximal compact subgroup $SU(2) \times SU(2) \times U(1)$ and a basis vector in the iso-Poincaré subgroup $E(3,1)$, thereby joining the two different approaches mentioned in the introduction. All fourteen different classes of unitary irreducible representations are studied, and we find that two classes are reducible with respect to the "lightlike" representations of $E(3,1)$, two classes are reducible with respect to the "timelike" representations, eight classes are reducible with respect to the "spacelike" representations, and two classes contain both "timelike" and "spacelike" representations of $E(3,1)$. The typical matrix elements involve, at worst, Whittaker's functions of the second kind and are given explicitly. The general matrix elements are obtainable by simple algebraic manipulations involving repeated differentiations of the typical matrix elements.

Section 8 ends with a very brief summary of our results. Since this paper at times relies heavily on the notations used and results obtained in Papers I and II, the readers are advised to consult those papers for more details.[5]

2. THE ISO-POINCARÉ SUBGROUP $E(3,1)$

The group $SU(2,2)$ and the reduction with respect to its maximal compact subgroup $SU(2) \times SU(2) \times U(1)$ have been studied.[5] The language and notations used there are peculiar to unitary groups. The canonical generators are A_μ^ν, $\mu, \nu = 1, 2, 3, 4$, and the commutation relations are

$$[A_\alpha^\beta, A_\gamma^\delta] = \delta_\gamma^\beta A_\alpha^\delta - \delta_\alpha^\delta A_\gamma^\beta, \quad \alpha, \beta, \gamma, \delta = 1, 2, 3, 4. \tag{2.1}$$

Since $SU(2,2)$ is the covering group of $O(4,2)$, just as $SU(4)$ is the covering group of $O(6)$, we may also use the language and notations that are peculiar to orthogonal groups. The generators of $O(4,2)$ are L_a^b, $a, b = 0, 1, 2, 3, 5, 6$; they are rotations or boosts in the a-b space. The canonical commutation relations are

$$[L_{ab}, L_{cd}] = i(g_{ac}L_{bd} - g_{ad}L_{bc} - g_{bc}L_{ad} + g_{bd}L_{ac}), \quad a, b, c, d = 0, 1, 2, 3, 5, 6. \tag{2.2}$$

The metric is so chosen that $g_{11} = g_{22} = g_{33} = g_{55} = -g_{00} = -g_{66} = 1$. In Appendix A the relationship between L_{ab}, A_α^β, and the 15 Dirac γ-matrices are written down. For the purpose of reducing $SU(2,2)$ with respect to an iso-Poincaré subgroup, the L_{ab} are much more convenient to use. We single out the iso-Minkowski space 1, 2, 5, 6, and rewrite

$$\mathfrak{L}_{\alpha\beta} \equiv L_{\alpha\beta}, \quad \alpha, \beta = 1, 2, 5, 6. \tag{2.3}$$

Next, we introduce

$$\mathfrak{T}_\alpha \equiv L_{0\alpha} + L_{3\alpha}, \tag{2.4}$$

$$\mathfrak{R}_\alpha \equiv L_{0\alpha} - L_{3\alpha}. \tag{2.5}$$

The remaining generator is L_{03}, which we leave untouched.

The generators of the iso-Poincaré subgroup are $\mathfrak{L}_{\alpha\beta}$ and $\mathfrak{T}_\alpha$. The commutation relations are the familiar ones, and are obtained from Eq. (2.2):

$$\begin{aligned} &[\mathfrak{T}_\alpha, \mathfrak{T}_\beta] = 0, \\ &[\mathfrak{L}_{\mu\nu}, \mathfrak{T}_\alpha] = i(g_{\mu\alpha}\mathfrak{T}_\nu - g_{\nu\alpha}\mathfrak{T}_\mu), \\ &[\mathfrak{L}_{\mu\nu}, \mathfrak{L}_{\alpha\beta}] = i(g_{\mu\alpha}\mathfrak{L}_{\nu\beta} - g_{\mu\beta}\mathfrak{L}_{\nu\alpha} - g_{\nu\alpha}\mathfrak{L}_{\mu\beta} + g_{\nu\beta}\mathfrak{L}_{\mu\alpha}). \end{aligned} \tag{2.6}$$

The $\mathfrak{R}_\alpha$ transform like a 4-vector under $\mathfrak{L}_{\mu\nu}$,

$$[\mathfrak{L}_{\mu\nu}, \mathfrak{R}_\alpha] = i(g_{\mu\alpha}\mathfrak{R}_\nu - g_{\nu\alpha}\mathfrak{R}_\mu), \tag{2.7}$$

and commute among themselves,

$$[\mathfrak{R}_\alpha, \mathfrak{R}_\beta] = 0. \tag{2.8}$$

However,

$$[\mathfrak{T}_\alpha, \mathfrak{R}_\beta] = -2i(\mathfrak{L}_{\alpha\beta} + g_{\alpha\beta}L_{03}). \tag{2.9}$$

Finally, we have the commutation relations between L_{03} and $\mathfrak{L}_{\alpha\beta}$, $\mathfrak{T}_\alpha$, and $\mathfrak{R}_\alpha$,

$$[L_{03}, L_{\alpha\beta}] = 0, \tag{2.10}$$

$$[L_{03}, \mathfrak{T}_\alpha] = -i\mathfrak{T}_\alpha, \tag{2.11}$$

$$[L_{03}, \mathfrak{R}_\alpha] = i\mathfrak{R}_\alpha. \tag{2.12}$$

Reduction of the Iso-Poincaré Group with Respect to Euclidean Subgroups

1. *The $E(2)$ Subgroup*

We start with the Euclidean subgroup in two dimensions. The three generators are $\mathfrak{T}_1$, $\mathfrak{T}_2$, and $\mathfrak{L}_{12}$. We introduce the following notation,

$$\mathfrak{T}_\pm \equiv \mathfrak{T}_1 \pm i\mathfrak{T}_2, \tag{2.13}$$

$$\mathfrak{L}_5 \equiv \mathfrak{L}_{12}, \tag{2.14}$$

and obtain

$$[\mathfrak{T}_+, \mathfrak{T}_-] = 0, \tag{2.15}$$

$$[\mathfrak{L}_5, \mathfrak{T}_\pm] = \pm\mathfrak{T}_\pm. \tag{2.16}$$

The Casimir operator of $E(2)$ is $\mathfrak{T}_+\mathfrak{T}_-$, which commutes with all three generators. Now let $|\epsilon, m\rangle$ be simultaneous eigenfunctions of $\mathfrak{T}_+\mathfrak{T}_-$ and $\mathfrak{L}_5$; then $|\epsilon, m\rangle$ form a canonical basis which span a unitary

irreducible representation of $E(2)$:

$$\mathfrak{T}_+\mathfrak{T}_- |\epsilon, m\rangle = \epsilon^2 |\epsilon, m\rangle, \quad \epsilon^2 \geq 0, \tag{2.17}$$

$$\mathfrak{L}_5 |\epsilon, m\rangle = m |\epsilon, m\rangle, \quad m = 0, \pm\tfrac{1}{2}, \pm 1, \cdots, \tag{2.18}$$

$$\mathfrak{T}_\pm |\epsilon, m\rangle = \epsilon |\epsilon, m \pm 1\rangle. \tag{2.19}$$

Since $\mathfrak{T}_1$ and $\mathfrak{T}_2$ are the "translation" operators and $\mathfrak{L}_5$ is the rotation operator in the $(1, 2)$ plane, ϵ may be interpreted as the magnitude of "momentum" in the $(1, 2)$ plane.

2. The $E(3)$ Subgroup

Next, we deal with the Euclidean subgroup in three dimensions. The six generators are $\mathfrak{T}_1$, $\mathfrak{T}_2$, $\mathfrak{T}_5$, $\mathfrak{L}_{12}$, $\mathfrak{L}_{25}$, and $\mathfrak{L}_{51}$. The Casimir operators are $\mathfrak{P}^2$ and $\mathfrak{L} \cdot \mathfrak{T}$, where we have introduced

$$\mathfrak{L}_1 \equiv \mathfrak{L}_{25}, \quad \mathfrak{L}_2 \equiv \mathfrak{L}_{51}, \quad \mathfrak{P}^2 = \mathfrak{T}_1^2 + \mathfrak{T}_2^2 + \mathfrak{T}_5^2$$

and

$$\mathfrak{L}_\pm \equiv \mathfrak{L}_1 \pm i\mathfrak{L}_2, \quad \mathfrak{L} \cdot \mathfrak{T} = \mathfrak{L}_1\mathfrak{T}_1 + \mathfrak{L}_2\mathfrak{T}_2 + \mathfrak{L}_5\mathfrak{T}_5 \tag{2.20}$$

$$[\mathfrak{P}^2, \mathfrak{T}_i] = [\mathfrak{P}^2, \mathfrak{L}_i] = [\mathfrak{L} \cdot \mathfrak{T}, \mathfrak{T}_i] = [\mathfrak{L} \cdot \mathfrak{T}, \mathfrak{L}_i] = 0, \quad i = 1, 2, 5. \tag{2.21}$$

Now let $|\xi, \epsilon; t, m\rangle$ be a canonical basis for $E(3)$, with

$$\mathfrak{P}^2 |\xi, \epsilon; t, m\rangle = \xi^2 |\xi, \epsilon; t, m\rangle, \quad \xi^2 \geq 0, \tag{2.22}$$

$$\mathfrak{L} \cdot \mathfrak{T} |\xi, \epsilon; t, m\rangle = t\xi |\xi, \epsilon; t, m\rangle, \quad t = 0, \pm\tfrac{1}{2}, \pm 1, \cdots. \tag{2.23}$$

Here, we may interpret ξ as the magnitude of "3-momentum" and t as "helicity." $\mathfrak{T}$ and $\mathfrak{L}$ are interpreted as "momentum" and "angular momentum" operators, respectively.

From Eqs. (2.17) and (2.22) we have

$$\mathfrak{T}_5 |\xi, \epsilon; t, m\rangle = (\xi^2 - \epsilon^2)^{\frac{1}{2}} |\xi, \epsilon; t, m\rangle, \tag{2.24}$$

and from Eqs. (2.23) and (2.19) we have

$$\tfrac{1}{2}\epsilon\mathfrak{L}_+ |\xi, \epsilon; t, m-1\rangle + \tfrac{1}{2}\epsilon\mathfrak{L}_- |\xi, \epsilon; t, m+1\rangle = [t\xi - m(\xi^2 - \epsilon^2)^{\frac{1}{2}}] |\xi, \epsilon; t, m\rangle. \tag{2.25}$$

Let

$$\mathfrak{L}_+ |\xi, \epsilon; t, m\rangle = \frac{1}{\epsilon}\left(-f\frac{\partial}{\partial\epsilon} + c + t\xi - m(\xi^2 - \epsilon^2)^{\frac{1}{2}}\right) \times |\xi, \epsilon; t, m+1\rangle \tag{2.26}$$

and

$$\mathfrak{L}_- |\xi, \epsilon; t, m\rangle = \frac{1}{\epsilon}\left(f\frac{\partial}{\partial\epsilon} - c + t\xi - m(\xi^2 - \epsilon^2)^{\frac{1}{2}}\right) \times |\xi, \epsilon; t, m-1\rangle, \tag{2.27}$$

where f and c are functions of ξ and ϵ but independent of t and m. Equations (2.26) and (2.27) are the most general expressions compatible with the commutation relations

$$[\mathfrak{L}_+, \mathfrak{L}_-] = 2\mathfrak{L}_5, \tag{2.28}$$

$$[\mathfrak{L}_5, \mathfrak{L}_\pm] = \pm\mathfrak{L}_\pm, \tag{2.29}$$

and at the same time satisfy Eq. (2.25). From $[\mathfrak{L}_+, \mathfrak{T}_+] = 0$ we may determine f:

$$f = \epsilon(\xi^2 - \epsilon^2)^{\frac{1}{2}}. \tag{2.30}$$

Therefore,

$$\mathfrak{L}_+ |\xi, \epsilon; t, m\rangle = \left(-(\xi^2 - \epsilon^2)^{\frac{1}{2}}\frac{\partial}{\partial\epsilon} + \frac{c}{\epsilon} + t\frac{\xi}{\epsilon} - m\frac{1}{\epsilon}(\xi^2 - \epsilon^2)^{\frac{1}{2}}\right) \times |\xi, \epsilon; t, m+1\rangle,$$

$$\mathfrak{L}_- |\xi, \epsilon; t, m\rangle = \left((\xi^2 - \epsilon^2)^{\frac{1}{2}}\frac{\partial}{\partial\epsilon} - \frac{c}{\epsilon} + t\frac{\xi}{\epsilon} - m\frac{1}{\epsilon}(\xi^2 - \epsilon^2)^{\frac{1}{2}}\right) \times |\xi, \epsilon; t, m-1\rangle. \tag{2.31}$$

In order to determine the function c, we consider unitary representations. Let $f(\xi, \epsilon; t, m) \equiv |\xi, \epsilon; t, m\rangle$; then for unitary representations we have

$$(f(\xi', \epsilon'; t', m+1), \mathfrak{L}_+ f(\xi, \epsilon; t, m)) = (\mathfrak{L}_- f(\xi', \epsilon'; t', m+1), f(\xi, \epsilon; t, m)). \tag{2.32}$$

We normalize the states as follows:

$$\langle \xi', \epsilon'; t', m' \mid \xi, \epsilon; t, m\rangle = \delta(\xi'^2 - \xi^2)\delta(\epsilon'^2 - \epsilon^2)\delta_{t't}\delta_{m'm}. \tag{2.33}$$

Substituting Eqs. (2.31) into Eq. (2.32) and using Eq. (2.33), we get

$$\left[-(\xi^2 - \epsilon^2)^{\frac{1}{2}}\frac{\partial}{\partial\epsilon} + \frac{1}{\epsilon}\left(c + \frac{(\xi^2 - \epsilon^2)^{\frac{1}{2}}}{2}\right)\right]\delta(\epsilon'^2 - \epsilon^2) = \left[(\xi^2 - \epsilon'^2)^{\frac{1}{2}}\frac{\partial}{\partial\epsilon'} - \frac{1}{\epsilon'}\left(c' + \frac{(\xi^2 - \epsilon'^2)^{\frac{1}{2}}}{2}\right)\right]\delta(\epsilon'^2 - \epsilon^2). \tag{2.34}$$

Upon integration with respect to ϵ^2 from 0 to ξ^2, we finally have

$$c = \tfrac{1}{2}\epsilon^2(\xi^2 - \epsilon^2)^{-\frac{1}{2}} - (\xi^2 - \epsilon^2)^{\frac{1}{2}}. \tag{2.35}$$

Equations (2.31) now become

$$\mathfrak{L}_+ |\xi, \epsilon; t, m\rangle = \left(-(\xi^2 - \epsilon^2)^{\frac{1}{2}}\frac{\partial}{\partial\epsilon} + \frac{\epsilon(\xi^2 - \epsilon^2)^{-\frac{1}{2}}}{2} + t\frac{\xi}{\epsilon} - (m+1)\frac{1}{\epsilon}(\xi^2 - \epsilon^2)^{\frac{1}{2}}\right) \times |\xi, \epsilon; t, m+1\rangle,$$

$$\mathfrak{L}_- |\xi, \epsilon; t, m\rangle = \left((\xi^2 - \epsilon^2)^{\frac{1}{2}}\frac{\partial}{\partial\epsilon} - \frac{\epsilon(\xi^2 - \epsilon^2)^{-\frac{1}{2}}}{2} + t\frac{\xi}{\epsilon} - (m-1)\frac{1}{\epsilon}(\xi^2 - \epsilon^2)^{\frac{1}{2}}\right) \times |\xi, \epsilon; t, m-1\rangle. \tag{2.36}$$

We have presented the derivation of Eqs. (2.36) in detail since in later discussions the same kind of approach will be used frequently and the operators involved will be of greater complexity.

3. *The $E(3, 1)$ Subgroup*

Now we are ready to discuss the iso-Poincaré subgroup, which we denote by $E(3, 1)$. The ten generators are $\mathfrak{P}_\alpha$ and $\mathfrak{L}_{\alpha\beta}$, $\alpha, \beta = 1, 2, 5, 6$. The two Casimir operators are $\mathfrak{P}_\mu\mathfrak{P}^\mu$ and $W_\mu W^\mu$, where $W_\mu = \frac{1}{2}\epsilon_{\mu\nu\alpha\beta}\mathfrak{P}^\nu\mathfrak{L}^{\alpha\beta}$:

$$[\mathfrak{P}_\mu\mathfrak{P}^\mu, \mathfrak{P}_\alpha] = [\mathfrak{P}_\mu\mathfrak{P}^\mu, \mathfrak{L}_{\alpha\beta}] = [W_\mu W^\mu, \mathfrak{P}_\alpha] = [W_\mu W^\mu, \mathfrak{L}_{\alpha\beta}] = 0. \quad (2.37)$$

Let $|\eta, \xi, \epsilon; s, t, m\rangle$ be a canonical basis for $E(3, 1)$, with

$$\mathfrak{P}_\mu\mathfrak{P}^\mu |\eta, \xi, \epsilon; s, t, m\rangle = -\eta^2 |\eta, \xi, \epsilon; s, t, m\rangle, \quad (2.38)$$

$$W_\mu W^\mu |\eta, \xi, \epsilon; s, t, m\rangle = \eta^2 s(s+1) |\eta, \xi, \epsilon; s, t, m\rangle, \quad s = 0, \tfrac{1}{2}, 1, \cdots. \quad (2.39)$$

Here, we may interpret η as the "mass" and s as the "spin." For $\eta^2 > 0$, we have "timelike" representations, for $\eta^2 = 0$, "lightlike" representations, and, for $\eta^2 < 0$, "spacelike" representations. These three different types of representations must be treated separately.

3. "TIMELIKE" UNITARY IRREDUCIBLE REPRESENTATIONS OF $E(3, 1)$, REDUCTION WITH RESPECT TO $E(3)$ AND $E(2)$

For "timelike" unitary irreducible representations of $E(3, 1)$, we have

$$\mathfrak{P}_\mu\mathfrak{P}^\mu |\eta, \xi, \epsilon; s, t, m\rangle = -\eta^2 |\eta, \xi, \epsilon; s, t, m\rangle, \quad (3.1)$$

$$W_\mu W^\mu |\eta, \xi, \epsilon; s, t, m\rangle = \eta^2 s(s+1) |\eta, \xi, \epsilon; s, t, m\rangle, \quad (3.2)$$

where

$$\eta^2 > 0, \quad \xi^2 \geq \epsilon^2 \geq 0,$$

$$s = 0, \tfrac{1}{2}, 1, \tfrac{3}{2}, \cdots, \quad t, m = 0, \pm\tfrac{1}{2}, \pm 1, \cdots,$$

$$\mathfrak{P}_\pm |\eta, \xi, \epsilon; s, t, m\rangle = \epsilon |\eta, \xi, \epsilon; s, t, m \pm 1\rangle, \quad (3.3)$$

$$\mathfrak{P}_5 |\eta, \xi, \epsilon; s, t, m\rangle = (\xi^2 - \epsilon^2)^{\frac{1}{2}} |\eta, \xi, \epsilon; s, t, m\rangle, \quad (3.4)$$

$$\mathfrak{P}_6 |\eta, \xi, \epsilon; s, t, m\rangle = (\xi^2 + \eta^2)^{\frac{1}{2}} |\eta, \xi, \epsilon; s, t, m\rangle, \quad (3.5)$$

$$\mathfrak{L}_5 |\eta, \xi, \epsilon; s, t, m\rangle = m |\eta, \xi, \epsilon; s, t, m\rangle, \quad (3.6)$$

$$\mathfrak{L}_+ |\eta, \xi, \epsilon; s, t, m\rangle = \left(-(\xi^2 - \epsilon^2)^{\frac{1}{2}}\frac{\partial}{\partial\epsilon} + \frac{\epsilon(\xi^2 - \epsilon^2)^{-\frac{1}{2}}}{2} + t\frac{\xi}{\epsilon} - (m+1)\frac{1}{\epsilon}(\xi^2 - \epsilon^2)^{\frac{1}{2}}\right) |\eta, \xi, \epsilon; s, t, m+1\rangle, \quad (3.7)$$

$$\mathfrak{L}_- |\eta, \xi, \epsilon; s, t, m\rangle = \left((\xi^2 - \epsilon^2)^{\frac{1}{2}}\frac{\partial}{\partial\epsilon} - \frac{\epsilon(\xi^2 - \epsilon^2)^{-\frac{1}{2}}}{2} + t\frac{\xi}{\epsilon} - (m-1)\frac{1}{\epsilon}(\xi^2 - \epsilon^2)^{\frac{1}{2}}\right) |\eta, \xi, \epsilon; s, t, m-1\rangle. \quad (3.8)$$

Our next task is to find irreducible representations for $\mathfrak{L}_{16}$, $\mathfrak{L}_{26}$, and $\mathfrak{L}_{56}$, the "boost" operators as differential operators in the function space $|\eta, \xi, \epsilon; s, t, m\rangle$. The procedure is exactly the same as that following Eq. (2.25), only much more complicated. The details are relegated to Appendix B. Here we present our results:

$$\mathfrak{L}_{\pm 6} \equiv \mathfrak{L}_{16} \pm i\mathfrak{L}_{26},$$

$$\begin{aligned}
&\mathfrak{L}_{+6} |\eta, \xi, \epsilon; s, t, m\rangle \\
&= -i(\xi^2 + \eta^2)^{\frac{1}{2}}\left(\frac{\epsilon}{\xi}\frac{\partial}{\partial\xi} + \frac{\partial}{\partial\epsilon} + \frac{\epsilon(\xi^2 + \eta^2)^{-1}}{2} + (m+1)\frac{1}{\epsilon} - t\frac{1}{\xi\epsilon}(\xi^2 - \epsilon^2)^{\frac{1}{2}}\right) \\
&\quad \times |\eta, \xi, \epsilon; s, t, m+1\rangle \\
&\quad - \frac{i}{2}\frac{\eta}{\xi^2}[\xi + (\xi^2 - \epsilon^2)^{\frac{1}{2}}][(s-t)(s+t+1)]^{\frac{1}{2}} \\
&\quad \times |\eta, \xi, \epsilon; s, t+1, m+1\rangle \\
&\quad - \frac{i}{2}\frac{\eta}{\xi^2}[\xi - (\xi^2 - \epsilon^2)^{\frac{1}{2}}][(s+t)(s-t+1)]^{\frac{1}{2}} \\
&\quad \times |\eta, \xi, \epsilon; s, t-1, m+1\rangle, \quad (3.9)
\end{aligned}$$

$$\begin{aligned}
&\mathfrak{L}_{-6} |\eta, \xi, \epsilon; s, t, m\rangle \\
&= -i(\xi^2 + \eta^2)^{\frac{1}{2}}\left(\frac{\epsilon}{\xi}\frac{\partial}{\partial\xi} + \frac{\partial}{\partial\epsilon} + \frac{\epsilon(\xi^2 + \eta^2)^{-1}}{2} - (m-1)\frac{1}{\epsilon} + t\frac{1}{\xi\epsilon}(\xi^2 - \epsilon^2)^{\frac{1}{2}}\right) \\
&\quad \times |\eta, \xi, \epsilon; s, t, m-1\rangle \\
&\quad + \frac{i}{2}\frac{\eta}{\xi^2}[\xi - (\xi^2 - \epsilon^2)^{\frac{1}{2}}][(s-t)(s+t+1)]^{\frac{1}{2}} \\
&\quad \times |\eta, \xi, \epsilon; s, t+1, m-1\rangle \\
&\quad + \frac{i}{2}\frac{\eta}{\xi^2}[\xi + (\xi^2 - \epsilon^2)^{\frac{1}{2}}][(s+t)(s-t+1)]^{\frac{1}{2}} \\
&\quad \times |\eta, \xi, \epsilon; s, t-1, m-1\rangle, \quad (3.10)
\end{aligned}$$

$$\begin{aligned}
&\mathfrak{L}_{56} |\eta, \xi, \epsilon; s, t, m\rangle \\
&= -i\left((\xi^2 + \eta^2)^{\frac{1}{2}}(\xi^2 - \epsilon^2)^{\frac{1}{2}}\frac{1}{\xi}\frac{\partial}{\partial\xi} + \frac{(\xi^2 - \epsilon^2)^{\frac{1}{2}}(\xi^2 + \eta^2)^{-\frac{1}{2}}}{2} + \frac{(\xi^2 + \eta^2)^{\frac{1}{2}}(\xi^2 - \epsilon^2)^{-\frac{1}{2}}}{2}\right) \\
&\quad \times |\eta, \xi, \epsilon; s, t, m\rangle \\
&\quad + \frac{i}{2}\frac{\epsilon\eta}{\xi^2}[(s-t)(s+t+1)]^{\frac{1}{2}} |\eta, \xi, \epsilon; s, t+1, m\rangle \\
&\quad - \frac{i}{2}\frac{\epsilon\eta}{\xi^2}[(s+t)(s-t+1)]^{\frac{1}{2}} |\eta, \xi, \epsilon; s, t-1, m\rangle. \quad (3.11)
\end{aligned}$$

Here $|\eta, \xi, \epsilon; s, t, m\rangle$ are basis vectors for a unitary irreducible representation of $E(3, 1)$, and

$$(f(\eta', \xi', \epsilon'; s', t', m), \mathfrak{L}_{56} f(\eta, \xi, \epsilon; s, t, m)) = (\mathfrak{L}_{56} f(\eta', \xi', \epsilon'; s', t', m), f(\eta, \xi, \epsilon; s, t, m)). \tag{3.12}$$

The basis vectors are normalized as follows:

$$\langle \eta', \xi', \epsilon'; s', t', m' \mid \eta, \xi, \epsilon; s, t, m\rangle = \delta(\eta'^2 - \eta^2)\delta(\xi'^2 - \xi^2)\delta(\epsilon'^2 - \epsilon^2)\delta_{s's}\delta_{t't}\delta_{m'm}. \tag{3.13}$$

From Eqs. (3.9)–(3.11) we see that $\Delta t = 0, \pm 1$ and, for given s, $t = -s, -s+1, \cdots, s$. It can be checked explicitly that Eqs. (3.3)–(3.11) satisfy the commutation relations Eq. (2.6) in the function space $|\eta, \xi, \epsilon; s, t, m\rangle$.

For completeness we here record the representations of W_μ:

$$\begin{aligned} W_+ |\eta, \xi, \epsilon; s, t, m\rangle &= -(\xi^2 + \eta^2)^{\frac{1}{2}}(\epsilon/\xi)t\, |\eta, \xi, \epsilon; s, t, m+1\rangle \\ &\quad - \tfrac{1}{2}(\eta/\xi)[\xi + (\xi^2 - \epsilon^2)^{\frac{1}{2}}][(s-t)(s+t+1)]^{\frac{1}{2}} \\ &\quad \times |\eta, \xi, \epsilon; s, t+1, m+1\rangle \\ &\quad + \tfrac{1}{2}(\eta/\xi)[\xi - (\xi^2 - \epsilon^2)^{\frac{1}{2}}][(s+t)(s-t+1)]^{\frac{1}{2}} \\ &\quad \times |\eta, \xi, \epsilon; s, t-1, m+1\rangle, \end{aligned} \tag{3.14}$$

$$\begin{aligned} W_- |\eta, \xi, \epsilon; s, t, m\rangle &= -(\xi^2 + \eta^2)^{\frac{1}{2}}(\epsilon/\xi)t\, |\eta, \xi, \epsilon; s, t, m-1\rangle \\ &\quad + \tfrac{1}{2}(\eta/\xi)[\xi - (\xi^2 - \epsilon^2)^{\frac{1}{2}}][(s-t)(s+t+1)]^{\frac{1}{2}} \\ &\quad \times |\eta, \xi, \epsilon; s, t+1, m-1\rangle \\ &\quad - \tfrac{1}{2}(\eta/\xi)[\xi + (\xi^2 - \epsilon^2)^{\frac{1}{2}}][(s+t)(s-t+1)]^{\frac{1}{2}} \\ &\quad \times |\eta, \xi, \epsilon; s, t-1, m-1\rangle, \end{aligned} \tag{3.15}$$

$$\begin{aligned} W_5 |\eta, \xi, \epsilon; s, t, m\rangle &= -\xi^{-1}(\xi^2 + \eta^2)^{\frac{1}{2}}(\xi^2 - \epsilon^2)^{\frac{1}{2}} t\, |\eta, \xi, \epsilon; s, t, m\rangle \\ &\quad + \tfrac{1}{2}(\epsilon\eta/\xi)[(s-t)(s+t+1)]^{\frac{1}{2}} |\eta, \xi, \epsilon; s, t+1, m\rangle \\ &\quad + \tfrac{1}{2}(\epsilon\eta/\xi)[(s+t)(s-t+1)]^{\frac{1}{2}} |\eta, \xi, \epsilon; s, t-1, m\rangle, \end{aligned} \tag{3.16}$$

$$W_6 |\eta, \xi, \epsilon; s, t, m\rangle = -\mathfrak{L} \cdot \mathfrak{P} |\eta, \xi, \epsilon; s, t, m\rangle = -t\xi |\eta, \xi, \epsilon; s, t, m\rangle. \tag{3.17}$$

4. REDUCTION OF $SU(2, 2)$ WITH RESPECT TO $E(3, 1)$ ("TIMELIKE" REGION)

Besides the ten generators of $E(3, 1)$ we have the remaining five generators L_{03} and $\mathcal{R}_\alpha$ of $SU(2, 2)$. The study of L_{03} is particularly simple. From Eqs. (2.10) and (2.11),

$$[L_{03}, \mathfrak{P}_\mu] = -i\mathfrak{P}_\mu, \tag{4.1}$$

$$[L_{03}, \mathfrak{L}_{\mu\nu}] = 0, \tag{4.2}$$

we have

$$e^{i\zeta L_{03}} \mathfrak{P}_\mu e^{-i\zeta L_{03}} = e^{\zeta}\mathfrak{P}_\mu, \tag{4.3}$$

$$e^{i\zeta L_{03}} \mathfrak{L}_{\mu\nu} e^{-i\zeta L_{03}} = \mathfrak{L}_{\mu\nu}, \tag{4.4}$$

where $e^{i\zeta L_{03}}$ is a finite boost. Therefore,

$$e^{i\zeta L_{03}} \mathfrak{P}_\mu \mathfrak{P}^\mu e^{-i\zeta L_{03}} = e^{2\zeta}\mathfrak{P}_\mu \mathfrak{P}^\mu, \tag{4.5}$$

$$e^{i\zeta L_{03}} W_\mu W^\mu e^{-i\zeta L_{03}} = e^{2\zeta} W_\mu W^\mu. \tag{4.6}$$

Hence,

$$e^{i\zeta L_{03}} |\eta, \xi, \epsilon; s, t, m\rangle = N(\zeta)\, |\eta e^{-\zeta}, \xi e^{-\zeta}, \epsilon e^{-\zeta}; s, t, m\rangle, \tag{4.7}$$

where

$$N(\zeta) = e^{-3\zeta} \tag{4.8}$$

is the result of normalization equation (3.13). Equation (4.7) leads directly to

$$L_{03} |\eta, \xi, \epsilon; s, t, m\rangle = i\left(\eta \frac{\partial}{\partial \eta} + \xi \frac{\partial}{\partial \xi} + \epsilon \frac{\partial}{\partial \epsilon} + 3\right) |\eta, \xi, \epsilon; s, t, m\rangle. \tag{4.9}$$

Next we tackle the very complicated problem of finding the representations for $\mathcal{R}_\alpha$. The procedure is again similar to that following Eq. (2.25); we let

$$\begin{aligned} \mathcal{R}_5 |\eta, \xi, \epsilon; s, t, m\rangle &= \sum_k \sum_l \left(A_{kl} \frac{\partial^2}{\partial \eta^2} + B_{kl} \frac{\partial^2}{\partial \xi^2} + C_{kl} \frac{\partial^2}{\partial \epsilon^2} \right. \\ &\quad + D_{kl} \frac{\partial}{\partial \eta} \frac{\partial}{\partial \xi} + E_{kl} \frac{\partial}{\partial \eta} \frac{\partial}{\partial \epsilon} + F_{kl} \frac{\partial}{\partial \xi} \frac{\partial}{\partial \epsilon} \\ &\quad \left. + a_{kl} \frac{\partial}{\partial \eta} + b_{kl} \frac{\partial}{\partial \xi} + c_{kl} \frac{\partial}{\partial \epsilon} + d_{kl}(m) \right) \\ &\quad \times |\eta, \xi, \epsilon; s + k, t + l, m\rangle, \end{aligned} \tag{4.10}$$

with $k = 0, \pm 1$ and $l = 0, \pm 1$. It is our task to determine the 90 functions $A_{kl}, \cdots, d_{kl}(m)$, with each one a function of η, ξ, ϵ, and s, t, m. It is a formidable and tedious job, and in Appendix C we shall give an outline of our method. Here, we shall present our

final results:

$$
\begin{aligned}
&\mathcal{R}_5\,|\eta,\xi,\epsilon;s,t,m\rangle\\
&=\Big[(\xi^2-\epsilon^2)^{\frac12}\Big(-\frac{\partial^2}{\partial\eta^2}+\frac{\partial^2}{\partial\xi^2}-\frac{\partial^2}{\partial\epsilon^2}+2\frac{\eta}{\xi}\frac{\partial}{\eta\partial}\frac{\partial}{\partial\xi}\Big)+[\eta(\xi^2-\epsilon^2)^{-\frac12}-\eta^{-1}(\xi^2-\epsilon^2)^{\frac12}]\frac{\partial}{\partial\eta}\\
&\quad+\xi^{-1}[4(\xi^2-\epsilon^2)^{\frac12}+\epsilon^2(\xi^2-\epsilon^2)^{-\frac12}]\frac{\partial}{\partial\xi}+[\epsilon(\xi^2-\epsilon^2)^{-\frac12}-\epsilon^{-1}(\xi^2-\epsilon^2)^{\frac12}]\frac{\partial}{\partial\epsilon}\\
&\quad+m^2\epsilon^{-2}(\xi^2-\epsilon^2)^{\frac12}-2mt\xi\epsilon^{-2}+t^2(\xi^2-\epsilon^2)^{\frac12}(\eta^{-2}+2\xi^{-2}+\epsilon^{-2})-s(s+1)\xi^{-2}(\xi^2-\epsilon^2)^{\frac12}\\
&\quad-2t\eta^{-2}\xi^{-1}(\xi^2-\epsilon^2)^{\frac12}(\xi^2+\eta^2)^{\frac12}\beta(s)-\tfrac14(\xi^2-\epsilon^2)^{\frac12}(\xi^2+\eta^2)^{-1}+\tfrac94(\xi^2-\epsilon^2)^{-\frac12}+\eta^{-2}(\xi^2-\epsilon^2)^{\frac12}\alpha(s)\Big]\\
&\quad\times|\eta,\xi,\epsilon;s,t,m\rangle\\
&\quad+[(s-t)(s+t+1)]^{\frac12}\epsilon\xi^{-1}\Big[-\xi^{-1}(\xi^2+\eta^2)^{\frac12}\Big(\frac{\partial}{\partial\eta}+\eta^{-1}(t+1)+\frac{\eta(\xi^2+\eta^2)^{-1}}{2}\Big)+\eta^{-1}\beta(s)\Big]\\
&\quad\times|\eta,\xi,\epsilon;s,t+1,m\rangle\\
&\quad+[(s+t)(s-t+1)]^{\frac12}\epsilon\xi^{-1}\Big[\xi^{-1}(\xi^2+\eta^2)^{\frac12}\Big(\frac{\partial}{\partial\eta}-\eta^{-1}(t-1)+\frac{\eta(\xi^2+\eta^2)^{-1}}{2}\Big)+\eta^{-1}\beta(s)\Big]\\
&\quad\times|\eta,\xi,\epsilon;s,t-1,m\rangle\\
&\quad+[(s+t+1)(s+t+2)]^{\frac12}\epsilon\xi^{-1}\eta^{-1}\gamma(s)\,|\eta,\xi,\epsilon;s+1,t+1,m\rangle\\
&\quad+[(s+t+1)(s-t+1)]^{\frac12}2\xi^{-1}\eta^{-2}(\xi^2-\epsilon^2)^{\frac12}(\xi^2+\eta^2)^{\frac12}\gamma(s)\,|\eta,\xi,\epsilon;s+1,t,m\rangle\\
&\quad-[(s-t+1)(s-t+2)]^{\frac12}\epsilon\xi^{-1}\eta^{-1}\gamma(s)\,|\eta,\xi,\epsilon;s+1,t-1,m\rangle\\
&\quad-[(s-t-1)(s-t)]^{\frac12}\epsilon\xi^{-1}\eta^{-1}\gamma(s-1)\,|\eta,\xi,\epsilon;s-1,t+1,m\rangle\\
&\quad+[(s+t)(s-t)]^{\frac12}2\xi^{-1}\eta^{-2}(\xi^2-\epsilon^2)^{\frac12}(\xi^2+\eta^2)^{\frac12}\gamma(s-1)\,|\eta,\xi,\epsilon;s-1,t,m\rangle\\
&\quad+[(s+t-1)(s+t)]^{\frac12}\epsilon\xi^{-1}\eta^{-1}\gamma(s-1)\,|\eta,\xi,\epsilon;s-1,t-1,m\rangle,
\end{aligned}
\tag{4.11}
$$

where $\alpha(s)$, $\beta(s)$, and $\gamma(s)$ are functions of s only and depend on the eigenvalues of the Casimir operators C_2, C_3, and C_4 of $SU(2,2)$.

The expressions for $\mathcal{R}_+$, $\mathcal{R}_-$, and $\mathcal{R}_6$ can be similarly written down:

$$
\begin{aligned}
&\mathcal{R}_+|\eta,\xi,\epsilon;s,t,m\rangle\\
&=\Big[\epsilon\Big(-\frac{\partial^2}{\partial\eta^2}+\frac{\partial^2}{\partial\xi^2}+\frac{\partial^2}{\partial\epsilon^2}+2\frac{\eta}{\xi}\frac{\partial}{\partial\eta}\frac{\partial}{\partial\xi}+2\frac{\xi}{\epsilon}\frac{\partial}{\partial\epsilon}\frac{\partial}{\partial\xi}+2\frac{\eta}{\epsilon}\frac{\partial}{\partial\epsilon}\frac{\partial}{\partial\eta}\Big)\\
&\quad+\frac{\eta}{\epsilon}\Big(2(m+1)-\frac{\epsilon^2}{\eta^2}-2t\frac{1}{\xi}(\xi^2-\epsilon^2)^{\frac12}\Big)\frac{\partial}{\partial\eta}\\
&\quad+\frac{\xi}{\epsilon}\Big(2(m+1)+3\frac{\epsilon^2}{\xi^2}-2t\frac{1}{\xi}(\xi^2-\epsilon^2)^{\frac12}\Big)\frac{\partial}{\partial\xi}+(2m+7)\frac{\partial}{\partial\epsilon}\\
&\quad+m^2\epsilon^{-1}+6m\epsilon^{-1}+t^2\epsilon(\eta^{-2}+2\xi^{-2}-\epsilon^{-2})-s(s+1)\epsilon\xi^{-2}-t\epsilon\xi^{-1}(\xi^2-\epsilon^2)^{-\frac12}\\
&\quad-4t\xi^{-1}\epsilon^{-1}(\xi^2-\epsilon^2)^{\frac12}-2t\epsilon\xi^{-1}\eta^{-2}(\xi^2+\eta^2)^{\frac12}\beta(s)+\tfrac52\epsilon^{-1}+\tfrac52\xi^2\epsilon^{-1}(\xi^2-\epsilon^2)^{-1}\\
&\quad-\tfrac14\epsilon(\xi^2+\eta^2)^{-1}-\tfrac94\epsilon(\xi^2-\epsilon^2)^{-1}+\epsilon\eta^{-2}\alpha(s)\Big]\,|\eta,\xi,\epsilon;s,t,m+1\rangle\\
&\quad+[(s-t)(s+t+1)]^{\frac12}\xi^{-1}[\xi+(\xi^2-\epsilon^2)^{\frac12}]\\
&\quad\times\Big[\xi^{-1}(\xi^2+\eta^2)^{\frac12}\Big(\frac{\partial}{\partial\eta}+\eta^{-1}(t+1)+\frac{\eta(\xi^2+\eta^2)^{-1}}{2}\Big)-\eta^{-1}\beta(s)\Big]\,|\eta,\xi,\epsilon;s,t+1,m+1\rangle\\
&\quad+[(s+t)(s-t+1)]^{\frac12}\xi^{-1}[\xi-(\xi^2-\epsilon^2)^{\frac12}]\\
&\quad\times\Big[\xi^{-1}(\xi^2+\eta^2)^{\frac12}\Big(\frac{\partial}{\partial\eta}-\eta^{-1}(t-1)+\frac{\eta(\xi^2+\eta^2)^{-1}}{2}\Big)+\eta^{-1}\beta(s)\Big]\,|\eta,\xi,\epsilon;s,t-1,m+1\rangle\\
&\quad-[(s+t+1)(s+t+2)]^{\frac12}\xi^{-1}\eta^{-1}[\xi+(\xi^2-\epsilon^2)^{\frac12}]\gamma(s)\,|\eta,\xi,\epsilon;s+1,t+1,m+1\rangle\\
&\quad+[(s+t+1)(s-t+1)]^{\frac12}2\epsilon\xi^{-1}\eta^{-2}(\xi^2+\eta^2)^{\frac12}\gamma(s)\,|\eta,\xi,\epsilon;s+1,t,m+1\rangle\\
&\quad-[(s-t+1)(s-t+2)]^{\frac12}\xi^{-1}\eta^{-1}[\xi-(\xi^2-\epsilon^2)^{\frac12}]\gamma(s)\,|\eta,\xi,\epsilon;s+1,t-1,m+1\rangle\\
&\quad+[(s-t)(s-t-1)]^{\frac12}\xi^{-1}[\xi+(\xi^2-\epsilon^2)^{\frac12}]\gamma(s-1)\,|\eta,\xi,\epsilon;s-1,t+1,m+1\rangle\\
&\quad+[(s-t)(s+t)]^{\frac12}2\epsilon\xi^{-1}\eta^{-2}(\xi^2+\eta^2)^{\frac12}\gamma(s-1)\,|\eta,\xi,\epsilon;s-1,t,m+1\rangle\\
&\quad+[(s+t)(s+t-1)]^{\frac12}\xi^{-1}[\xi-(\xi^2-\epsilon^2)^{\frac12}]\gamma(s-1)\,|\eta,\xi,\epsilon;s-1,t-1,m+1\rangle,
\end{aligned}
\tag{4.12}
$$

$$
\begin{aligned}
&\mathcal{R}_- |\eta, \xi, \epsilon; s, t, m\rangle \\
&= \Big[\epsilon\Big(-\frac{\partial^2}{\partial\eta^2}+\frac{\partial^2}{\partial\xi^2}+\frac{\partial^2}{\partial\epsilon^2}+2\frac{\eta}{\xi}\frac{\partial}{\partial\eta}\frac{\partial}{\partial\xi}+2\frac{\xi}{\epsilon}\frac{\partial}{\partial\epsilon}\frac{\partial}{\partial\xi}+2\frac{\eta}{\epsilon}\frac{\partial}{\partial\epsilon}\frac{\partial}{\partial\eta}\Big) \\
&\quad+\frac{\eta}{\epsilon}\Big(-2(m-1)-\frac{\epsilon^2}{\eta^2}+2t\frac{1}{\xi}(\xi^2-\epsilon^2)^{\frac{1}{2}}\Big)\frac{\partial}{\partial\eta} \\
&\quad+\frac{\xi}{\epsilon}\Big(-2(m-1)+3\frac{\epsilon^2}{\xi^2}+2t\frac{1}{\xi}(\xi^2-\epsilon^2)^{\frac{1}{2}}\Big)\frac{\partial}{\partial\xi}-(2m-7)\frac{\partial}{\partial\epsilon} \\
&\quad+m^2\epsilon^{-1}-6m\epsilon^{-1}+t^2\epsilon(\eta^{-2}+2\xi^{-2}-\epsilon^{-2})-s(s+1)\epsilon\xi^{-2}+t\epsilon\xi^{-1}(\xi^2-\epsilon^2)^{-\frac{1}{2}} \\
&\quad+4t\xi^{-1}\epsilon^{-1}(\xi^2-\epsilon^2)^{\frac{1}{2}}-2t\epsilon\xi^{-1}\eta^{-2}(\xi^2+\eta^2)^{\frac{1}{2}}\beta(s)+\tfrac{5}{2}\epsilon^{-1}+\tfrac{5}{2}\xi^2\epsilon^{-1}(\xi^2-\epsilon^2)^{-1} \\
&\quad-\tfrac{1}{4}\epsilon(\xi^2+\eta^2)^{-1}-\tfrac{9}{4}\epsilon(\xi^2-\epsilon^2)^{-1}+\epsilon\eta^{-2}\alpha(s)\Big]|\eta, \xi, \epsilon; s, t, m-1\rangle \\
&\quad-[(s-t)(s+t+1)]^{\frac{1}{2}}\xi^{-1}[\xi-(\xi^2-\epsilon^2)^{\frac{1}{2}}] \\
&\quad\times\Big[\xi^{-1}(\xi^2+\eta^2)^{\frac{1}{2}}\Big(\frac{\partial}{\partial\eta}+\eta^{-1}(t+1)+\frac{\eta(\xi^2+\eta^2)^{-1}}{2}\Big)-\eta^{-1}\beta(s)\Big]|\eta, \xi, \epsilon; s, t+1, m-1\rangle \\
&\quad-[(s+t)(s-t+1)]^{\frac{1}{2}}\xi^{-1}[\xi+(\xi^2-\epsilon^2)^{\frac{1}{2}}] \\
&\quad\times\Big[\xi^{-1}(\xi^2+\eta^2)^{\frac{1}{2}}\Big(\frac{\partial}{\partial\eta}-\eta^{-1}(t-1)+\frac{\eta(\xi^2+\eta^2)^{-1}}{2}\Big)+\eta^{-1}\beta(s)\Big]|\eta, \xi, \epsilon; s, t-1, m-1\rangle \\
&\quad+[(s+t+1)(s+t+2)]^{\frac{1}{2}}\xi^{-1}\eta^{-1}[\xi-(\xi^2-\epsilon^2)^{\frac{1}{2}}]\gamma(s)|\eta, \xi, \epsilon; s+1, t+1, m-1\rangle \\
&\quad+[(s+t+1)(s-t+1)]^{\frac{1}{2}}2\epsilon\xi^{-1}\eta^{-2}(\xi^2+\eta^2)^{\frac{1}{2}}\gamma(s)|\eta, \xi, \epsilon; s+1, t, m-1\rangle \\
&\quad+[(s-t+1)(s-t+2)]^{\frac{1}{2}}\xi^{-1}\eta^{-1}[\xi+(\xi^2-\epsilon^2)^{\frac{1}{2}}]\gamma(s)|\eta, \xi, \epsilon; s+1, t-1, m-1\rangle \\
&\quad-[(s-t)(s-t-1)]^{\frac{1}{2}}\xi^{-1}\eta^{-1}[\xi-(\xi^2-\epsilon^2)^{\frac{1}{2}}]\gamma(s-1)|\eta, \xi, \epsilon; s-1, t+1, m-1\rangle \\
&\quad+[(s-t)(s+t)]^{\frac{1}{2}}2\epsilon\xi^{-1}\eta^{-2}(\xi^2+\eta^2)^{\frac{1}{2}}\gamma(s-1)|\eta, \xi, \epsilon; s-1, t, m-1\rangle \\
&\quad-[(s+t)(s+t-1)]^{\frac{1}{2}}\xi^{-1}\eta^{-1}[\xi+(\xi^2-\epsilon^2)^{\frac{1}{2}}]\gamma(s-1)|\eta, \xi, \epsilon; s-1, t-1, m-1\rangle, \qquad (4.13)
\end{aligned}
$$

$$
\begin{aligned}
&\mathcal{R}_6 |\eta, \xi, \epsilon; s, t, m\rangle \\
&= \Big[-(\xi^2+\eta^2)^{\frac{1}{2}}\Big(\frac{\partial^2}{\partial\eta^2}+\frac{\partial^2}{\partial\xi^2}+\frac{\partial^2}{\partial\epsilon^2}+2\frac{\epsilon}{\xi}\frac{\partial}{\partial\xi}\frac{\partial}{\partial\epsilon}\Big)-[\eta(\xi^2+\eta^2)^{-\frac{1}{2}}+\eta^{-1}(\xi^2+\eta^2)^{\frac{1}{2}}]\frac{\partial}{\partial\eta} \\
&\quad-\xi^{-1}[4(\xi^2+\eta^2)^{\frac{1}{2}}-\eta^2(\xi^2+\eta^2)^{-\frac{1}{2}}]\frac{\partial}{\partial\xi}-[\epsilon(\xi^2+\eta^2)^{-\frac{1}{2}}+\epsilon^{-1}(\xi^2+\eta^2)^{\frac{1}{2}}]\frac{\partial}{\partial\epsilon} \\
&\quad+m^2\epsilon^{-2}(\xi^2+\eta^2)^{\frac{1}{2}}-2mt\epsilon^{-2}\xi^{-1}(\xi^2+\eta^2)^{\frac{1}{2}}(\xi^2-\epsilon^2)^{\frac{1}{2}}+t^2(\xi^2+\eta^2)^{\frac{1}{2}}(\eta^{-2}-2\xi^{-2}+\epsilon^{-2}) \\
&\quad+s(s+1)\xi^{-2}(\xi^2+\eta^2)^{\frac{1}{2}}-2t\xi\eta^{-2}\beta(s)+\tfrac{1}{4}(\xi^2+\eta^2)^{\frac{1}{2}}(\xi^2-\epsilon^2)^{-1}-\tfrac{9}{4}(\xi^2+\eta^2)^{-\frac{1}{2}}+\eta^{-2}(\xi^2+\eta^2)^{\frac{1}{2}}\alpha(s)\Big] \\
&\quad\times|\eta, \xi, \epsilon; s, t, m\rangle \\
&\quad+[(s-t)(s+t+1)]^{\frac{1}{2}}\frac{\eta}{\xi}\Big[-\xi^{-1}(\xi^2-\epsilon^2)^{\frac{1}{2}}\Big(\frac{\partial}{\partial\epsilon}+(t+1)\epsilon^{-1}-\frac{\epsilon(\xi^2-\epsilon^2)^{-1}}{2}\Big)+m\epsilon^{-1}\Big] \\
&\quad\times|\eta, \xi, \epsilon; s, t+1, m\rangle \\
&\quad+[(s+t)(s-t+1)]^{\frac{1}{2}}\frac{\eta}{\xi}\Big[\xi^{-1}(\xi^2-\epsilon^2)^{\frac{1}{2}}\Big(\frac{\partial}{\partial\epsilon}-(t-1)\epsilon^{-1}-\frac{\epsilon(\xi^2-\epsilon^2)^{-1}}{2}\Big)+m\epsilon^{-1}\Big] \\
&\quad\times|\eta, \xi, \epsilon; s, t-1, m\rangle \\
&\quad+2[(s+t+1)(s-t+1)]^{\frac{1}{2}}\xi\eta^{-2}\gamma(s)|\eta, \xi, \epsilon; s+1, t, m\rangle \\
&\quad+2[(s-t)(s+t)]^{\frac{1}{2}}\xi\eta^{-2}\gamma(s-1)|\eta, \xi, \epsilon; s-1, t, m\rangle. \qquad (4.14)
\end{aligned}
$$

Equation (4.14) for $\mathcal{R}_6$ is relatively simpler since it is the sum of five states. Equations (4.11)–(4.13) all contain nine states where $\Delta s = 0, \pm 1$ and $\Delta t = 0, \pm 1$. The reader is invited to check that Eqs. (4.11)–(4.14) satisfy the commutation relations (2.8) in the function space $|\eta, \xi, \epsilon; s, t, m\rangle$, if

$$(2s+3)\gamma^2(s)-(2s-1)\gamma^2(s-1)=\beta^2(s)-\alpha(s). \qquad (4.15)$$

We turn our attention next to the determination of the three functions $\alpha(s)$, $\beta(s)$, and $\gamma(s)$. But, first, we have to consider the three Casimir operators C_2, C_3, and C_4 of $SU(2, 2)$. For unitary irreducible representations of $SU(2, 2)$, we have

$$C_2 |c_2, c_3, c_4; \eta, \xi, \epsilon; s, t, m\rangle = c_2 |c_2, c_3, c_4; \eta, \xi, \epsilon; s, t, m\rangle, \quad (4.16)$$

$$C_3 |c_2, c_3, c_4; \eta, \xi, \epsilon; s, t, m\rangle = c_3 |c_2, c_3, c_4; \eta, \xi, \epsilon; s, t, m\rangle, \quad (4.17)$$

$$C_4 |c_2, c_3, c_4; \eta, \xi, \epsilon; s, t, m\rangle = c_4 |c_2, c_3, c_4; \eta, \xi, \epsilon; s, t, m\rangle, \quad (4.18)$$

where c_2, c_3, and c_4 are the eigenvalues of C_2, C_3, and C_4.

A. The Casimir Operator C_2

The quadratic Casimir operator C_2 has been defined in Ref. 5, and in the $O(4, 2)$ and $E(3, 1)$ language can be written as

$$C_2 = \tfrac{1}{2}L_{ab}L^{ab} = \tfrac{1}{2}\mathfrak{L}_{\mu\nu}\mathfrak{L}^{\mu\nu} + 4iL_{03} - L_{03}^2 - \mathcal{R}_\mu\mathcal{P}^\mu. \quad (4.19)$$

Substituting Eqs. (3.3)–(3.11), (4.9), and (4.11)–(4.14) into Eqs. (4.19) and (4.16), we get, after tremendous cancellation,

$$[\alpha(s) + 2s(s + 1) - 4] |c_2, c_3, c_4; \eta, \xi, \epsilon; s, t, m\rangle = c_2 |c_2, c_3, c_4; \eta, \xi, \epsilon; s, t, m\rangle$$

or

$$\alpha(s) = c_2 - 2s(s + 1) + 4. \quad (4.20)$$

B. The Casimir Operator C_3

The cubic Casimir operator C_3 can be written as

$$C_3 = -\tfrac{1}{48}\epsilon^{abcdef}L_{ab}L_{cd}L_{ef} = -\tfrac{1}{4}(W^\mu\mathcal{R}_\mu + \mathcal{R}_\mu W^\mu) + \tfrac{1}{8}\epsilon^{\mu\nu\alpha\beta}L_{03}\mathfrak{L}_{\mu\nu}\mathfrak{L}_{\alpha\beta}, \quad (4.21)$$

where ϵ^{abcdef} is a totally antisymmetric rank-6 tensor. Again, substituting Eqs. (3.3)–(3.11), (4.9), and (4.11)–(4.14) into Eqs. (4.21) and (4.17), we get

$$\beta(s) = -c_3/s(s + 1). \quad (4.22)$$

C. The Casimir Operator C_4

The quartic Casimir operator C_4 is particularly complicated:

$$\begin{aligned} 4C_4 &= L_{ab}L^{bc}L_{cd}L^{da} - C_2^2 - 8C_2 \\ &= \mathcal{R}_\mu\mathcal{R}^\mu\mathcal{P}_\nu\mathcal{P}^\nu + 4\mathcal{R}^\alpha\mathfrak{L}_{\alpha\beta}\mathfrak{L}^{\beta\gamma}\mathcal{P}_\gamma - 4\mathcal{R}^\mu\mathfrak{L}_{\mu\nu}\mathcal{P}^\nu(L_{03} - 6i) \\ &\quad + \tfrac{3}{4}(\mathfrak{L}_{\mu\nu}\mathfrak{L}^{\mu\nu})^2 + \tfrac{1}{16}(\epsilon^{\mu\nu\alpha\beta}\mathfrak{L}_{\mu\nu}\mathfrak{L}_{\alpha\beta})^2 \\ &\quad + \mathfrak{L}_{\mu\nu}\mathfrak{L}^{\mu\nu}(L_{03}^2 - 8iL_{03} - C_2 - 22) \\ &\quad - L_{03}^4 + 16iL_{03}^3 + 80L_{03}^2 \\ &\quad - 128iL_{03} + 36C_2 + 16iC_2L_{03} - 2C_2L_{03}^2. \end{aligned} \quad (4.23)$$

We again substitute Eqs. (3.3)–(3.11), (4.9), and (4.11)–(4.14) into Eqs. (4.23) and (4.18), and get

$$\begin{aligned} &\{c_2^2 - 4s(s + 1)c_2 - 4[c_3^2/s(s + 1)] \\ &\quad + 4s^2(s + 1)^2 - 8s(s + 1) + 4c_2 \\ &\quad - 4(s + 1)(2s + 3)\gamma^2(s) - 4s(2s - 1)\gamma^2(s - 1)\} \\ &\quad \times |c_2, c_3, c_4; \eta, \xi, \epsilon; s, t, m\rangle \\ &= 4c_4 |c_2, c_3, c_4; \eta, \xi, \epsilon; s, t, m\rangle. \end{aligned} \quad (4.24)$$

Substituting Eqs. (4.20) and (4.22) into Eq. (4.15), we obtain

$$\gamma^2(s) = \frac{1}{4}\left(s^2 + 2s - c_2 - \frac{15}{4}\right) + \frac{a}{(2s + 1)(2s + 3)} - \frac{c_3^2}{(2s + 1)(2s + 3)(s + 1)^2}, \quad (4.25)$$

where a is a constant independent of s. Now, Eq. (4.24) is satisfied, provided that

$$a = \tfrac{1}{4}(c_2 + \tfrac{5}{2})(c_2 + \tfrac{9}{2}) - c_4. \quad (4.26)$$

In passing, we remark that, in the process of obtaining Eqs. (4.20), (4.22), and (4.25), several thousand terms appear at the beginning of the calculation. They involve functions of η, ξ, ϵ, s, t, and m and also differential operators up to fourth order in η, ξ, and ϵ. Most of them cancel and leave us only the results of Eqs. (4.20), (4.22), and (4.24), which involve s only. This calculation, therefore, also serves as an independent check that Eqs. (4.11)–(4.14) are right and the arithmetic has been performed correctly.

D. The Allowed Values of C_2, C_3, and C_4

The eigenvalues c_2, c_3, and c_4 of the three Casimir operators can best be determined with the help of the maximal compact subgroup $SU(2) \times SU(2) \times U(1)$ of $SU(2, 2)$, which was studied earlier.

From Eqs. (4.25)–(4.26), we have

$$(2s + 1)(2s + 3)(s + 1)^2\gamma^2(s) = (s + 1)^6 - (c_2 + 5)(s + 1)^4 + [\tfrac{1}{4}(c_2 + 4)^2 - c_4](s + 1)^2 - c_3^2, \quad (4.27)$$

which we write as

$$(2s + 1)(2s + 3)(s + 1)^2\gamma^2(s) = [(s + 1)^2 - A^2][(s + 1)^2 - B^2][(s + 1)^2 - C^2]. \quad (4.28)$$

Comparing Eq. (4.28) with Eq. (4.27), we find the following parametrization:

$$C_2 = A^2 + B^2 + C^2 - 5, \quad (4.29)$$

$$C_3 = ABC, \quad (4.30)$$

$$C_4 = \tfrac{1}{4}[A^2 + B^2 + C^2 - 1]^2 - [A^2B^2 + B^2C^2 + C^2A^2]. \quad (4.31)$$

For degenerate unitary irreducible representations, c_2, c_3, and c_4 depend on two parameters only, and, by comparing Eqs. (4.29)–(4.31) to Eqs. (1.9)–(1.11) of Paper II, we see that C is not independent,

$$C = B + 1, \tag{4.32}$$

and that there are 14 types of degenerate unitary irreducible representations with allowed values for A and B given in Sec. 4 of that paper. [However, we caution the reader that *not* all 14 degenerate unitary irreducible representations of $SU(2, 2)$ are reducible with respect to the "timelike" unitary irreducible representations of the iso-Poincaré subgroup $E(3, 1)$. Some of them are reducible with respect to the "spacelike," others with respect to the "lightlike" UIR's of $E(3, 1)$. This problem will be treated in Sec. 7.]

For nondegenerate unitary irreducible representations of $SU(2, 2)$, the three parameters A, B, and C are linearly independent of each other. One example we have is from Paper I, which has a discussion on the discrete series $D^{\pm}$. (The existence of $D^{\pm}$ is the consequence of a general theorem due to Harish-Chandra.) If we consider the D^- series, we have, from Eqs. (6.8)–(6.10) of Paper I,

$$C_2 = 2J_m(J_m + 1) + 2K_m(K_m + 1) + \Lambda_m(\Lambda_m + 4), \tag{4.33}$$

$$C_3 = -(\Lambda_m + 2)(J_m - K_m)(J_m + K_m + 1), \tag{4.34}$$

$$C_4 = \tfrac{1}{4}[(\Lambda_m + 2)^2 - 4J_m(J_m + 1)] \times [(\Lambda_m + 2)^2 - 4K_m(K_m + 1)] - (\Lambda_m + 2)^2, \tag{4.35}$$

which, upon comparison with Eqs. (4.29)–(4.31), give

$$A = \pm(\Lambda_m + 2) \quad (+, K_m > J_m; -, J_m > K_m), \tag{4.36}$$

$$B = |J_m - K_m|, \tag{4.37}$$

$$C = J_m + K_m + 1, \tag{4.38}$$

where J_m, $K_m = 0, \frac{1}{2}, 1, \frac{3}{2}, \cdots$ and $\Lambda_m = -J_m - K_m - S_m - 4$, with $S_m = -1, 0, 1, 2, 3, \cdots$ [see Eq. (6.15) of Paper I]. If either J_m or K_m equals 0, S_m may also be -2, besides the allowed values -1, 0, 1, 2, $\cdots$; then we are dealing with the degenerate UIR's of the D^- series, and $C = B + 1$.

At this stage we wish to emphasize that, in general, it is not easy to determine the allowed values of A, B, and C to obtain unitary irreducible representations of $SU(2, 2)$. The discussion we presented above is based completely on results from Papers I and II, where we studied the reduction $SU(2, 2) \supset SU(2) \times SU(2) \times U(1)$. The only general remark we can make concerning Eq. (4.28) is that at least one of the three parameters A, B, and C must be discrete, say $A = s_0$, $s_0 = 0, \frac{1}{2}, 1, \frac{3}{2}, \cdots$; then $\gamma^2(s_0 - 1) = 0$, and $s = s_0$ is the smallest "spin." The other two parameters B and C may be discrete or continuous; in fact, they may even be complex, subject only to the condition that Eq. (4.28) is real and positive.

5. REDUCTION OF $SU(2, 2)$ WITH RESPECT TO $E(3, 1)$ ("SPACELIKE" REGION)

For "spacelike" unitary irreducible representations of $E(3, 1)$, we have

$$\mathfrak{P}_\mu \mathfrak{P}^\mu |\omega, \xi, \epsilon; r, t, m\rangle = \omega^2 |\omega, \xi, \epsilon; r, t, m\rangle, \tag{5.1}$$

$$W_\mu W^\mu |\omega, \xi, \epsilon; r, t, m\rangle = -\omega^2 r(r - 1) |\omega, \xi, \epsilon; r, t, m\rangle, \tag{5.2}$$

where

$$\omega^2 > 0, \quad \xi^2 \geq \omega^2, \quad \xi^2 \geq \epsilon^2 \geq 0, \quad t, m = 0, \pm\tfrac{1}{2}, \pm 1, \cdots$$

Equations (5.1) and (5.2) are obtained from Eqs. (3.1) and (3.2) by replacing η by $i\omega$ and $s(s + 1)$ by $r(r - 1)$. In contrast to the "timelike" region, where the little group of $E(3, 1)$ is $SU(2) \sim O(3)$, in the "spacelike" region the little group of $E(3, 1)$ is $SU(1, 1) \sim O(2, 1)$. The eigenvalues of the Casimir operator of the little group $SU(2)$ are $s(s + 1)$ with $s = 0, \frac{1}{2}, 1, \cdots$; the eigenvalues of the Casimir operator of the little group $SU(1, 1)$ are $r(r - 1)$, which according to Bargmann can be divided into four types[6]:

(1) C_q^0, the even continuous series, $q = -r(r - 1) > 0$, $m = \pm$ integer,

(2) $C_q^{\frac{1}{2}}$, the odd continuous series, $q = -r(r - 1) > \frac{1}{4}$, $m = \pm$ half-integer,

(3) the D^+ discrete series, $r = \frac{1}{2}, 1, \frac{3}{2}, \cdots$, $m = r, r + 1, \cdots$, and

(4) the D^- discrete series, $r = \frac{1}{2}, 1, \frac{3}{2}, \cdots$, $m = -r, -r - 1, \cdots$.

Expressions for "spacelike" unitary irreducible representations of $E(3, 1)$ analogous to Eqs. (3.3)–(3.11) and Eqs. (4.11)–(4.14) can now be written down immediately by the replacement of η by $i\omega$ and s by $r - 1$. Instead of Eqs. (4.20), (4.22), and (4.28), we have now

$$\alpha(r) = c_2 - 2r(r - 1) + 4, \tag{5.3}$$

$$\beta(r) = -c_3/r(r - 1), \tag{5.4}$$

$$(2r - 1)(2r + 1)r^2\gamma^2(r) = (r^2 - A^2)(r^2 - B^2)(r^2 - C^2). \tag{5.5}$$

6. REDUCTION OF $SU(2,2)$ WITH RESPECT TO $E(3,1)$ ("LIGHTLIKE" REGIONS)

The unitary irreducible representations of $E(3,1)$ which have zero "mass" are divided into two classes: (1) those with discrete "spin" and (2) those with continuous "spin."

A. "Lightlike" Unitary Irreducible Representations of $E(3,1)$ with Discrete "Spin"

The states here are labeled by ξ, ϵ, t, and m only:

$$\mathfrak{T}_\mu \mathfrak{T}^\mu |\xi, \epsilon; t, m\rangle = 0, \tag{6.1}$$

$$W_\mu W^\mu |\xi, \epsilon; t, m\rangle = 0, \tag{6.2}$$

$$\mathfrak{T}_\pm |\xi, \epsilon; t, m\rangle = \epsilon |\xi, \epsilon; t, m \pm 1\rangle, \tag{6.3}$$

$$\mathfrak{T}_5 |\xi, \epsilon; t, m\rangle = (\xi^2 - \epsilon^2)^{\frac{1}{2}} |\xi, \epsilon; t, m\rangle, \tag{6.4}$$

$$\mathfrak{T}_6 |\xi, \epsilon; t, m\rangle = \xi |\xi, \epsilon; t, m\rangle, \tag{6.5}$$

$$\mathfrak{L}_5 |\xi, \epsilon; t, m\rangle = m |\xi, \epsilon; t, m\rangle, \tag{6.6}$$

$$\mathfrak{L}_+ |\xi, \epsilon; t, m\rangle = \left(-(\xi^2 - \epsilon^2)^{\frac{1}{2}} \frac{\partial}{\partial \epsilon} + \frac{\epsilon(\xi^2 - \epsilon^2)^{-\frac{1}{2}}}{2} + t\frac{\xi}{\epsilon} - (m+1)\frac{1}{\epsilon}(\xi^2 - \epsilon^2)^{\frac{1}{2}}\right) |\xi, \epsilon; t, m+1\rangle, \tag{6.7}$$

$$\mathfrak{L}_- |\xi, \epsilon; t, m\rangle = \left((\xi^2 - \epsilon^2)^{\frac{1}{2}} \frac{\partial}{\partial \epsilon} - \frac{\epsilon(\xi^2 - \epsilon^2)^{-\frac{1}{2}}}{2} + t\frac{\xi}{\epsilon} - (m-1)\frac{1}{\epsilon}(\xi^2 - \epsilon^2)^{\frac{1}{2}}\right) |\xi, \epsilon; t, m-1\rangle, \tag{6.8}$$

$$\mathfrak{L}_{+6} |\xi, \epsilon; t, m\rangle = -i\left(\epsilon \frac{\partial}{\partial \xi} + \xi \frac{\partial}{\partial \epsilon} + \frac{1}{2}\frac{\epsilon}{\xi} + (m+1)\frac{\xi}{\epsilon} - t\frac{1}{\epsilon}(\xi^2 - \epsilon^2)^{\frac{1}{2}}\right) |\xi, \epsilon; t, m+1\rangle, \tag{6.9}$$

$$\mathfrak{L}_{-6} |\xi, \epsilon; t, m\rangle = -i\left(\epsilon \frac{\partial}{\partial \xi} + \xi \frac{\partial}{\partial \epsilon} + \frac{1}{2}\frac{\epsilon}{\xi} - (m-1)\frac{\xi}{\epsilon} + t\frac{1}{\epsilon}(\xi^2 - \epsilon^2)^{\frac{1}{2}}\right) |\xi, \epsilon, t, m-1\rangle, \tag{6.10}$$

$$\mathfrak{L}_{56} |\xi, \epsilon; t, m\rangle = -i\left((\xi^2 - \epsilon^2)^{\frac{1}{2}} \frac{\partial}{\partial \xi} + \frac{1}{2}\frac{1}{\xi}(\xi^2 - \epsilon^2)^{\frac{1}{2}} + \frac{\xi(\xi^2 - \epsilon^2)^{-\frac{1}{2}}}{2}\right) |\xi, \epsilon, t, m\rangle, \tag{6.11}$$

$$W_\pm |\xi, \epsilon; t, m\rangle = -t\epsilon |\xi, \epsilon; t, m \pm 1\rangle, \tag{6.12}$$

$$W_5 |\xi, \epsilon; t, m\rangle = -t(\xi^2 - \epsilon^2)^{\frac{1}{2}} |\xi, \epsilon; t, m\rangle, \tag{6.13}$$

$$W_6 |\xi, \epsilon; t, m\rangle = -t\xi |\xi, \epsilon; t, m\rangle. \tag{6.14}$$

Equations (6.1)–(6.14) are obtained directly from Eqs. (3.1)–(3.11) and (3.14)–(3.17) by setting $\eta = 0$. From Eqs. (6.12)–(6.14) we see that W_μ behaves like $-t\mathfrak{T}_\mu$ when acting on $|\xi, \epsilon; t, m\rangle$. The basis vectors are now normalized differently,

$$\langle \xi', \epsilon'; t', m' \,|\, \xi, \epsilon; t, m\rangle = \delta(\xi'^2 - \xi^2)\delta(\epsilon'^2 - \epsilon^2)\delta_{t't}\delta_{m'm}, \tag{6.15}$$

with the result that

$$e^{i\zeta L_{03}} |\xi, \epsilon; t, m\rangle = e^{-2\zeta} |\xi e^{-\zeta}, \epsilon e^{-\zeta}; t, m\rangle \tag{6.16}$$

and

$$L_{03} |\xi, \epsilon; t, m\rangle = i\left(\xi \frac{\partial}{\partial \xi} + \epsilon \frac{\partial}{\partial \epsilon} + 2\right) |\xi, \epsilon; t, m\rangle. \tag{6.17}$$

Next, we find the representations for $\mathfrak{R}_\alpha$. The method is the same as in Appendix C, only much simpler this time:

$$\mathfrak{R}_5 |\xi, \epsilon; t, m\rangle = \Big[(\xi^2 - \epsilon^2)^{\frac{1}{2}}\left(\frac{\partial^2}{\partial \xi^2} - \frac{\partial^2}{\partial \epsilon^2}\right) + [\xi^{-1}(\xi^2 - \epsilon^2)^{\frac{1}{2}} + \xi(\xi^2 - \epsilon^2)^{-\frac{1}{2}}]\frac{\partial}{\partial \xi} + [\epsilon(\xi^2 - \epsilon^2)^{-\frac{1}{2}} - \epsilon^{-1}(\xi^2 - \epsilon^2)^{\frac{1}{2}}]\frac{\partial}{\partial \epsilon} + (\xi^2 - \epsilon^2)^{\frac{1}{2}}[(t^2 + m^2)\epsilon^{-2} - \tfrac{1}{4}\xi^{-2}] - 2mt\xi\epsilon^{-2} + (\xi^2 - \epsilon^2)^{-\frac{1}{2}}\delta(t)\Big] |\xi, \epsilon; t, m\rangle, \tag{6.18}$$

$$\mathfrak{R}_+ |\xi, \epsilon; t, m\rangle = \Big[\epsilon\left(\frac{\partial^2}{\partial \xi^2} + \frac{\partial^2}{\partial \epsilon^2} + 2\frac{\xi}{\epsilon}\frac{\partial}{\partial \epsilon}\frac{\partial}{\partial \xi}\right) + [\epsilon\xi^{-1} + 2(m+1)\xi\epsilon^{-1} - 2t\epsilon^{-1}(\xi^2 - \epsilon^2)^{\frac{1}{2}}]\frac{\partial}{\partial \xi} + (2m+5)\frac{\partial}{\partial \epsilon} - t^2\epsilon^{-1} - t\epsilon^{-1}[\xi(\xi^2 - \epsilon^2)^{-\frac{1}{2}} + \xi^{-1}(\xi^2 - \epsilon^2)^{\frac{1}{2}}] + \epsilon^{-1}(m^2 + 4m + \tfrac{3}{2}) + \tfrac{3}{2}\xi^2\epsilon^{-1}(\xi^2 - \epsilon^2)^{-1} - \tfrac{1}{4}\epsilon\xi^{-2} - \epsilon(\xi^2 - \epsilon^2)^{-1}\delta(t)\Big] |\xi, \epsilon; t, m+1\rangle, \tag{6.19}$$

$$\mathcal{R}_-|\xi,\epsilon;t,m\rangle=\Big[\epsilon\Big(\frac{\partial^2}{\partial\xi^2}+\frac{\partial^2}{\partial\epsilon^2}+2\frac{\xi}{\epsilon}\frac{\partial}{\partial\epsilon}\frac{\partial}{\partial\xi}\Big)+[\epsilon\xi^{-1}-2(m-1)\xi\epsilon^{-1}+2t\epsilon^{-1}(\xi^2-\epsilon^2)^{\frac12}]\frac{\partial}{\partial\xi}$$
$$-(2m-5)\frac{\partial}{\partial\epsilon}-t^2\epsilon^{-1}+t\epsilon^{-1}[\xi(\xi^2-\epsilon^2)^{-\frac12}+\xi^{-1}(\xi^2-\epsilon^2)^{\frac12}]+\epsilon^{-1}(m^2-4m+\tfrac32)$$
$$+\tfrac32\xi^2\epsilon^{-1}(\xi^2-\epsilon^2)^{-1}-\tfrac14\epsilon\xi^{-2}-\epsilon(\xi^2-\epsilon^2)^{-1}\delta(t)\Big]|\xi,\epsilon;t,m-1\rangle, \tag{6.20}$$

$$\mathcal{R}_6|\xi,\epsilon;t,m\rangle=\Big[-\xi\Big(\frac{\partial^2}{\partial\xi^2}+\frac{\partial^2}{\partial\epsilon^2}+2\frac{\epsilon}{\xi}\frac{\partial}{\partial\xi}\frac{\partial}{\partial\epsilon}\Big)-4\frac{\partial}{\partial\xi}-\Big(\frac{\xi}{\epsilon}+\frac{\epsilon}{\xi}\Big)\frac{\partial}{\partial\epsilon}+t^2\frac{\xi}{\epsilon^2}-2mt\epsilon^{-2}(\xi^2-\epsilon^2)^{\frac12}$$
$$+m^2\xi\epsilon^{-2}+\tfrac32\xi(\xi^2-\epsilon^2)^{-1}-\tfrac54\xi^{-1}-\xi(\xi^2-\epsilon^2)^{-1}\delta(t)\Big]|\xi,\epsilon;t,m\rangle, \tag{6.21}$$

where $\delta(t)$ is yet to be determined.

It is interesting to note that in Eqs. (6.18)–(6.21) the "helicity" t is unchanged, $\Delta t=0$, in contrast to Eqs. (4.11)–(4.14) where $\Delta t=0,\pm1$. Next, we apply Eqs. (4.19), (4.21), and (4.23) to the state $|\xi,\epsilon;t,m\rangle$, and find

$$C_2=3t^2-\tfrac12-2\delta(t), \tag{6.22a}$$
$$C_3=t[\delta(t)-t^2-\tfrac14], \tag{6.22b}$$
$$C_4=-\tfrac14\{3t^4-t^2[1+4\delta(t)]+12[\delta(t)-1]\}. \tag{6.22c}$$

Since C_2, C_3, and C_4 depend on two parameters t and $\delta(t)$ only, the unitary irreducible representations under study here are degenerate. When we compare Eqs. (6.22) with the results listed in Sec. IV of Paper II, we find that only Classes XI and XII (the exceptional degenerate discrete series $E^{\pm}$) are compatible, with $t=0,\ \pm\frac12,\ \pm1,\cdots$, and $\delta(t)=\frac54$. We note that Eqs. (6.22) now become

$$C_2=3(t^2-1), \tag{6.23a}$$
$$C_3=-t(t^2-1), \tag{6.23b}$$
$$C_4=-\tfrac34(t^2-1)^2 \tag{6.23c}$$

and are also in the form of Eqs. (4.29)–(4.31) with $A=-t\pm1$, $B=\pm t$, and $C=B+1$. [See also Eqs. (7.24).]

B. "Lightlike" Unitary Irreducible Representations of $E(3,1)$ with Continuous "Spin"

The states here are labeled by ζ, ξ, ϵ, t, and m:

$$\mathfrak{T}_\mu\mathfrak{T}^\mu|\zeta;\xi,\epsilon;t,m\rangle=0, \tag{6.24}$$
$$W_\mu W^\mu|\zeta;\xi,\epsilon;t,m\rangle=\zeta^2|\zeta;\xi,\epsilon;t,m\rangle, \tag{6.25}$$
$$\mathfrak{T}_\pm|\zeta;\xi,\epsilon;t,m\rangle=\epsilon|\zeta;\xi,\epsilon;t,m\pm1\rangle, \tag{6.26}$$
$$\mathfrak{T}_5|\zeta;\xi,\epsilon;t,m\rangle=(\xi^2-\epsilon^2)^{\frac12}|\zeta;\xi,\epsilon,t,m\rangle, \tag{6.27}$$
$$\mathfrak{T}_6|\zeta;\xi,\epsilon;t,m\rangle=\xi|\zeta;\xi,\epsilon;t,m\rangle, \tag{6.28}$$
$$\mathfrak{L}_5|\zeta;\xi,\epsilon;t,m\rangle=m|\zeta;\xi,\epsilon;t,m\rangle, \tag{6.29}$$

$$\mathfrak{L}_\pm|\zeta;\xi,\epsilon;t,m\rangle=\Big(\mp(\xi^2-\epsilon^2)^{\frac12}\frac{\partial}{\partial\epsilon}\pm\frac{\epsilon(\xi^2-\epsilon^2)^{-\frac12}}{2}+t\frac{\xi}{\epsilon}$$
$$-(m\pm1)\frac{1}{\epsilon}(\xi^2-\epsilon^2)^{\frac12}\Big)|\zeta;\xi,\epsilon;t,m\pm1\rangle, \tag{6.30}$$

$$\mathfrak{L}_{\pm6}|\zeta;\xi,\epsilon;t,m\rangle=-i\Big(\epsilon\frac{\partial}{\partial\xi}+\xi\frac{\partial}{\partial\epsilon}+\frac12\frac{\epsilon}{\xi}\pm(m\pm1)\frac{\xi}{\epsilon}$$
$$\mp t\frac{1}{\epsilon}(\xi^2-\epsilon^2)^{\frac12}\Big)|\zeta;\xi,\epsilon;t,m\pm1\rangle$$
$$\mp\tfrac12 i\xi^{-2}[\xi\pm(\xi^2-\epsilon^2)^{\frac12}]\zeta|\zeta;\xi,\epsilon;t+1,m\pm1\rangle$$
$$\mp\tfrac12 i\xi^{-2}[\xi\mp(\xi^2-\epsilon^2)^{\frac12}]\zeta|\zeta;\xi,\epsilon;t-1,m\pm1\rangle, \tag{6.31}$$

$$\mathfrak{L}_{56}|\zeta;\xi,\epsilon;t,m\rangle=i\Big((\xi^2-\epsilon^2)^{\frac12}\frac{\partial}{\partial\xi}+\frac12\frac{1}{\xi}(\xi^2-\epsilon^2)^{\frac12}$$
$$+\frac{\xi(\xi^2-\epsilon^2)^{-\frac12}}{2}\Big)|\zeta;\xi,\epsilon;t,m\rangle$$
$$+\tfrac12 i\epsilon\xi^{-2}\zeta|\zeta;\xi,\epsilon;t+1,m\rangle$$
$$-\tfrac12 i\epsilon\xi^{-2}\zeta|\zeta;\xi,\epsilon;t-1,m\rangle. \tag{6.32}$$

Equations (6.24)–(6.32) are obtained from Eqs. (3.1)–(3.11) by taking the limit $\eta=0$, but $\eta^2 s(s+1)=\zeta^2$. Equations (3.14)–(3.17) now become

$$W_\pm|\zeta;\xi,\epsilon;t,m\rangle=-\epsilon t|\zeta;\xi,\epsilon;t,m\pm1\rangle$$
$$\mp\tfrac12\xi^{-1}[\xi\pm(\xi^2-\epsilon^2)^{\frac12}]\zeta|\zeta;\xi,\epsilon;t+1,m\pm1\rangle$$
$$\pm\tfrac12\xi^{-1}[\xi\mp(\xi^2-\epsilon^2)^{\frac12}]\zeta|\zeta;\xi,\epsilon;t-1,m\pm1\rangle, \tag{6.33}$$

$$W_5|\zeta;\xi,\epsilon;t,m\rangle=-(\xi^2-\epsilon^2)^{\frac12}t|\zeta;\xi,\epsilon;t,m\rangle$$
$$+\tfrac12\epsilon\xi^{-1}\zeta|\zeta;\xi,\epsilon;t+1,m\rangle$$
$$+\tfrac12\epsilon\xi^{-1}\zeta|\zeta;\xi,\epsilon;t-1,m\rangle, \tag{6.34}$$

$$W_6|\zeta;\xi,\epsilon;t,m\rangle=-t\xi|\zeta;\xi,\epsilon;t,m\rangle. \tag{6.35}$$

From Eqs. (4.3)–(4.6), we again have

$$e^{i\alpha L_{03}}|\zeta;\xi,\epsilon;t,m\rangle = e^{-3\alpha}|\zeta e^{-\alpha};\xi e^{-\alpha},\epsilon e^{-\alpha};t,m\rangle \tag{6.36}$$

and

$$L_{03}|\zeta;\xi,\epsilon;t,m\rangle = i\left(\zeta\frac{\partial}{\partial\zeta}+\xi\frac{\partial}{\partial\xi}+\epsilon\frac{\partial}{\partial\epsilon}+3\right)|\zeta;\xi,\epsilon,t,m\rangle, \tag{6.37}$$

with the basis vectors normalized as follows:

$$\langle\zeta';\xi',\epsilon';t',m'\,|\,\zeta;\xi,\epsilon;t,m\rangle = \delta(\zeta'^2-\zeta^2)\delta(\xi'^2-\xi^2)\delta(\epsilon'^2-\epsilon^2)\delta_{t't}\delta_{m'm}. \tag{6.38}$$

We next turn our attention to the representation for $\mathcal{R}_\alpha$. Following the same procedure as in Appendix C, we write

$$\mathcal{R}_5|\zeta;\xi,\epsilon;t,m\rangle = \sum_l\left(A_l\frac{\partial^2}{\partial\zeta^2}+B_l\frac{\partial^2}{\partial\xi^2}+C_l\frac{\partial^2}{\partial\epsilon^2}+D_l\frac{\partial}{\partial\zeta}\frac{\partial}{\partial\xi}+E_l\frac{\partial}{\partial\zeta}\frac{\partial}{\partial\epsilon}+F_l\frac{\partial}{\partial\xi}\frac{\partial}{\partial\epsilon}+a_l\frac{\partial}{\partial\zeta}+b_l\frac{\partial}{\partial\xi}+c_l\frac{\partial}{\partial\epsilon}+d_l(m)\right)\times|\zeta;\xi,\epsilon;t+l,m\rangle \tag{6.39}$$

and try to determine the A_l, B_l, etc. However, it becomes immediately obvious that we *cannot* find a set of the A_l, B_l, etc., which satisfy the commutation relations (2.7), (2.9), and (2.12). We conclude, therefore, that, in the reduction of $SU(2,2)$ with respect to $E(3,1)$, the continuous "spin" representations of $E(3,1)$ do *not* contribute.[7]

C. Unitary Irreducible Representations of $E(3,1)$ with $\mathfrak{T}_\mu = 0$

When $\mathfrak{T}_\mu = 0$, $\mu = 1, 2, 5, 6$, the little group is the iso-Lorentz group $O(3,1)$. The states are now labeled by k_0, c, s, and m:

$$\mathfrak{T}_\mu\mathfrak{T}^\mu|k_0,c;s,m\rangle = 0, \tag{6.40}$$

$$W_\mu W^\mu|k_0,c;s,m\rangle = 0, \tag{6.41}$$

$$\mathfrak{L}_{\mu\nu}\mathfrak{L}^{\mu\nu}|k_0,c;s,m\rangle = 2(k_0^2+c^2-1)|k_0,c;s,m\rangle, \tag{6.42}$$

$$\epsilon_{\mu\nu\alpha\beta}\mathfrak{L}^{\mu\nu}\mathfrak{L}^{\alpha\beta}|k_0,c;s,m\rangle = -8ik_0c|k_0,c;s,m\rangle, \tag{6.43}$$

$$\mathfrak{T}_\mu|k_0,c;s,m\rangle = 0, \tag{6.44}$$

$$\mathfrak{L}_5|k_0,c;s,m\rangle = m|k_0,c;s,m\rangle, \tag{6.45}$$

$$\mathfrak{L}_\pm|k_0,c;s,m\rangle = [(s\mp m)(s\pm m+1)]^{\frac{1}{2}}\times|k_0,c;s,m\pm1\rangle, \tag{6.46}$$

$$\mathfrak{L}_{56}|k_0,c;s,m\rangle = [(s-m)(s+m)]^{\frac{1}{2}}\times C_s|k_0,c;s-1,m\rangle - mA_s|k_0,c;s,m\rangle - [(s+m+1)(s-m+1)]^{\frac{1}{2}}\times C_{s+1}|k_0,c;s+1,m\rangle, \tag{6.47}$$

$$\mathfrak{L}_{\pm6}|k_0,c;s,m\rangle = \pm[(s\mp m)(s\mp m-1)]^{\frac{1}{2}}\times C_s|k_0,c;s-1,m\pm1\rangle - [(s\mp m)(s\pm m+1)]^{\frac{1}{2}}\times A_s|k_0,c;s,m\pm1\rangle \pm[(s\pm m+1)(s\pm m+2)]^{\frac{1}{2}}\times C_{s+1}|k_0,c;s+1,m\pm1\rangle, \tag{6.48}$$

where $\mathfrak{L}_{\mu\nu}\mathfrak{L}^{\mu\nu}$ and $\epsilon_{\mu\nu\alpha\beta}\mathfrak{L}^{\mu\nu}\mathfrak{L}^{\alpha\beta}$ are the Casimir operators of the iso-Lorentz group $O(3,1)$, and[8]

$$A_s = ik_0c/s(s+1), \tag{6.49}$$

$$C_s = (i/s)[(s^2-k_0^2)(s^2-c^2)/(4s^2-1)]^{\frac{1}{2}}, \tag{6.50}$$

with either (1) c purely imaginary and k_0 half-integer or integer (the principal series) or (2) c a real number, $0\le c\le1$, and $k_0=0$ (the complementary series).

However, when we next study the representations of L_{03}, and $\mathcal{R}_\alpha$, we find immediately that *no* representations exist for $\mathcal{R}_\alpha$ which satisfy the commutation relations (2.7), (2.9), and (2.12). We conclude, therefore, that, in the reduction of $SU(2,2)$ with respect to $E(3,1)$, the null representation of $E(3,1)$ also does *not* contribute.

7. DETERMINATION OF MATRIX ELEMENTS OF $SU(2,2)$

In Secs. 4–6 we have studied the reduction of $SU(2,2)$ with respect to an iso-Poincaré subgroup $E(3,1)$. The 15 generators of $SU(2,2)$ have been expressed as differential operators in the Hilbert space spanned by the basis vectors $|\eta,\xi,\epsilon;s,t,m\rangle$, $|\omega,\xi,\epsilon;r,t,m\rangle$, or $|\xi,\epsilon;t,m\rangle$. Our next task is to determine the allowed eigenvalues of the three Casimir operators C_2, C_3, and C_4 and the matrix elements of $SU(2,2)$. In this connection, the maximal compact subgroup $SU(2)\times SU(2)\times U(1)$ plays an important role.

We recall briefly the notations used in Papers I and II. The basis vector is $|j,\mu;k,\nu;\lambda\rangle$, with

$$\mathbf{J}^2|j,\mu;k,\nu;\lambda\rangle = j(j+1)|j,\mu;k,\nu;\lambda\rangle, \tag{7.1}$$

$$J_3|j,\mu;k,\nu;\lambda\rangle = \mu|j,\mu;k,\nu;\lambda\rangle, \tag{7.2}$$

$$\mathbf{K}^2|j,\mu;k,\nu;\lambda\rangle = k(k+1)|j,\mu;k,\nu;\lambda\rangle, \tag{7.3}$$

$$K_3|j,\mu;k,\nu;\lambda\rangle = \nu|j,\mu;k,\nu;\lambda\rangle, \tag{7.4}$$

$$R_0|j,\mu;k,\nu;\lambda\rangle = \lambda|j,\mu;k,\nu;\lambda\rangle, \tag{7.5}$$

where $\mathbf{J}$, $\mathbf{K}$, and R_0 are the generators of $SU(2)\times SU(2)\times U(1)$.

A. The "Lightlike" Representations of $E(3,1)$

We start with the simpler "lightlike" representations of $E(3,1)$, treated in Sec. 6, since we know already that the exceptional degenerate discrete series $E^{\pm}$ of $SU(2,2)$, upon reduction with respect to $E(3,1)$, contain "lightlike" representations only.

1. UIR's in the E^+ Series

Consider one of the "ground states" of E^+,

$$|j = p_0, \mu = p_0; k = 0, \nu = 0; \lambda = p_0 + 1\rangle,$$

where $p_0 = 0, \frac{1}{2}, 1, \cdots$; then, from Appendix A, we have

$$\mathfrak{L}_+ |j = p_0, \mu = p_0; k = 0, \nu = 0; \lambda = p_0 + 1\rangle = -i(J_+ - K_+)|p_0, p_0; 0, 0; p_0 + 1\rangle = 0. \quad (7.6)$$

Since we are considering unitary representations, the matrix element

$$\langle p_0, p_0; 0, 0; p_0 + 1|\, \mathfrak{L}_- \,|\xi, \epsilon; t, m\rangle = 0. \quad (7.7)$$

Define

$$f(\xi, \epsilon; t) \equiv \langle p_0, p_0; 0, 0; p_0 + 1 \mid \xi, \epsilon; t, m = p_0\rangle; \quad (7.8)$$

then we have, using Eq. (6.8),

$$\left((\xi^2 - \epsilon^2)^{\frac{1}{2}} \frac{\partial}{\partial \epsilon} - \frac{\epsilon(\xi^2 - \epsilon^2)^{-\frac{1}{2}}}{2} + t\frac{\xi}{\epsilon} - p_0 \frac{1}{\epsilon}(\xi^2 - \epsilon^2)^{\frac{1}{2}}\right) \times f(\xi, \epsilon; t) = 0, \quad (7.9)$$

which can be solved easily:

$$f(\xi, \epsilon; t) = c(\xi, t)\epsilon^{p_0 - t}(\xi^2 - \epsilon^2)^{-\frac{1}{4}}[\xi + (\xi^2 - \epsilon^2)^{\frac{1}{2}}]^t, \quad (7.10)$$

with $c(\xi, t)$ still to be determined. Next we consider the relation

$$L_{35} = \tfrac{1}{2}(\mathfrak{I}_5 - \mathfrak{K}_5) = J_3 - K_3, \quad (7.11)$$

which leads to

$$\langle p_0, p_0; 0, 0; p_0 + 1|\, \mathfrak{K}_5 \,|\xi, \epsilon; t, m\rangle = [(\xi^2 - \epsilon^2)^{\frac{1}{2}} - 2p_0] \times \langle p_0, p_0; 0, 0; p_0 + 1 \mid \xi, \epsilon; t, m\rangle \quad (7.12)$$

Equation (6.18) now gives us

$$\Big[(\xi^2 - \epsilon^2)^{\frac{1}{2}}\left(\frac{\partial^2}{\partial \xi^2} - \frac{\partial^2}{\partial \epsilon^2}\right) + [\xi^{-1}(\xi^2 - \epsilon^2)^{\frac{1}{2}} + \xi(\xi^2 - \epsilon^2)^{-\frac{1}{2}}]\frac{\partial}{\delta \xi} + [\epsilon(\xi^2 - \epsilon^2)^{-\frac{1}{2}} - \epsilon^{-1}(\xi^2 - \epsilon^2)^{\frac{1}{2}}]\frac{\partial}{\partial \epsilon} + (\xi^2 - \epsilon^2)^{\frac{1}{2}}[t^2\epsilon^{-2} + p_0^2\epsilon^{-2} - \tfrac{1}{4}\xi^{-2}] - 2p_0 t\xi\epsilon^{-2} + (\xi^2 - \epsilon^2)^{-\frac{1}{2}}\delta(t) - (\xi^2 - \epsilon^2)^{\frac{1}{2}} + 2p_0\Big] f(\xi, \epsilon; t) = 0, \quad (7.13)$$

which, upon substituting Eq. (7.10) in it, reduces to

$$\Big((\xi^2 - \epsilon^2)^{\frac{1}{2}}\frac{d^2}{d\xi^2} + [2t + \xi^{-1}(\xi^2 - \epsilon^2)^{\frac{1}{2}}]\frac{d}{d\xi} + (\xi^2 - \epsilon^2)^{-\frac{1}{2}}[\delta(t) - \tfrac{5}{4}] - (\xi^2 - \epsilon^2)^{\frac{1}{2}} \times [1 + \tfrac{1}{4}\xi^{-2}] + t\xi^{-1} + 2p_0\Big) c(\xi, t) = 0. \quad (7.14)$$

Since $c(\xi, t)$ is independent of ϵ, Eq. (7.14) reduces down to three separate equations:

$$\left[\frac{d^2}{d\xi^2} + \frac{1}{\xi}\frac{d}{d\xi} - \left(1 + \frac{1}{4}\frac{1}{\xi^2}\right)\right] c(\xi, t) = 0, \quad (7.15)$$

$$\left(2t\frac{d}{d\xi} + t\frac{1}{\xi} + 2p_0\right) c(\xi, t) = 0, \quad (7.16)$$

$$\delta(t) = \tfrac{5}{4}. \quad (7.17)$$

Equation (7.15) is easily solved,

$$c(\xi, t) = c(t)e^{-\xi}\xi^{-\frac{1}{2}}, \quad (7.18)$$

and Eq. (7.16) gives

$$t = p_0, \quad (7.19)$$

which reduces Eq. (7.10) to

$$f(\xi, \epsilon; p_0) = c(p_0)e^{-\xi}\xi^{-\frac{1}{2}}(\xi^2 - \epsilon^2)^{-\frac{1}{4}}[\xi + (\xi^2 - \epsilon^2)^{\frac{1}{2}}]^{p_0}, \quad (7.20)$$

where $c(p_0)$ is a normalization constant.

Next we consider the relation

$$L_{06} = \tfrac{1}{2}(\mathfrak{I}_6 + \mathfrak{K}_6) = R_0, \quad (7.21)$$

which leads to

$$\langle p_0, p_0; 0, 0; p_0 + 1|\, \mathfrak{K}_6 \,|\xi, \epsilon; t, m\rangle = [2(p_0 + 1) - \xi]\langle p_0, p_0; 0, 0; p_0 + 1 \mid \xi, \epsilon; t, m\rangle. \quad (7.22)$$

Equation (6.22) now gives us

$$\Big[-\xi\left(\frac{\partial^2}{\partial \xi^2} + \frac{\partial^2}{\partial \epsilon^2} + 2\frac{\epsilon}{\xi}\frac{\partial}{\partial \xi}\frac{\partial}{\partial \epsilon}\right) - 4\frac{\partial}{\partial \xi} - \left(\frac{\xi}{\epsilon} + \frac{\epsilon}{\xi}\right)\frac{\partial}{\partial \epsilon} + t^2\xi\epsilon^{-2} - 2p_0 t\epsilon^{-2}(\xi^2 - \epsilon^2)^{\frac{1}{2}} + p_0^2\xi\epsilon^{-2} + \tfrac{3}{2}\xi(\xi^2 - \epsilon^2)^{-1} - \tfrac{5}{4}\xi^{-1} - \xi(\xi^2 - \epsilon^2)^{-1}\delta(t) + \xi - 2(p_0 + 1)\Big] f(\xi, \epsilon; t) = 0, \quad (7.23)$$

which is satisfied when we substitute Eqs. (7.17), (7.19), and (7.20) into it. The reason that we have to check with Eq. (7.23) is that both $|p_0, p_0; 0, 0; p_0 + 1\rangle$ (in the E^+ series) and $|p_0, p_0; 0, 0; -p_0 - 1\rangle$ (in the E^- series) lead to Eq. (7.20), but *only* the former also satisfies Eq. (7.23).

Equations (6.22) now become

$$c_2 = 3(p_0^2 - 1), \tag{7.24a}$$
$$c_3 = -p_0(p_0^2 - 1), \tag{7.24b}$$
$$c_4 = -\tfrac{3}{4}(p_0^2 - 1)^2, \tag{7.24c}$$

which are identical to Eqs. (6.23).

If, instead of $|j = p_0,\ \mu = p_0;\ k = 0,\ \nu = 0;\ \lambda = p_0 + 1\rangle$, we had chosen $|j = 0,\ \mu = 0;\ k = p_0,\ \nu = p_0;\ \lambda = p_0 + 1\rangle$, which belongs to another unitary irreducible representation in the E^+ series, we would have gotten again $\delta(t) = \frac{5}{4}$, but $t = -p_0$, with the result that

$$g(\xi, \epsilon; t) \equiv \langle 0, 0; p_0, p_0; p_0 + 1 \mid \xi, \epsilon; t, m = p_0\rangle$$
$$= c(p_0)e^{-\xi}\xi^{-\frac{1}{2}}(\xi^2 - \epsilon^2)^{-\frac{1}{4}}[\xi - (\xi^2 - \epsilon^2)^{\frac{1}{2}}]^{p_0} \tag{7.25}$$

and

$$c_2 = 3(p_0^2 - 1), \tag{7.26a}$$
$$c_3 = p_0(p_0^2 - 1), \tag{7.26b}$$
$$c_4 = -\tfrac{3}{4}(p_0^2 - 1)^2. \tag{7.26c}$$

2. *UIR's in the E^- Series*

Next we consider the unitary irreducible representations in the E^- series, which can be obtained from the E^+ series by replacing λ by $-\lambda$. We recall that

$$\mathfrak{T}_6 |\xi, \epsilon; t, m\rangle = \xi |\xi, \epsilon; t, m\rangle, \tag{6.5}$$

where we have chosen the "energy" to be positive. We could equally well have chosen the "energy" negative,

$$\mathfrak{T}_6 |\xi, \epsilon; t, m\rangle = -\xi |\xi, \epsilon; t, m\rangle, \tag{7.27}$$

with the result that we would have to replace ξ by $-\xi$ in Eqs. (6.7)–(6.22). We observe that Eqs. (7.9) and (7.13) are unchanged if we replace ξ by $-\xi$ and t by $-t$. Therefore, again

$$f'(\xi, \epsilon; p_0) \equiv \langle p_0, p_0; 0, 0; -p_0 - 1 \mid \xi, \epsilon; t, m = p_0\rangle$$
$$= c'(p_0)e^{-\xi}\xi^{-\frac{1}{2}}(\xi^2 - \epsilon^2)^{-\frac{1}{4}}[\xi + (\xi^2 - \epsilon^2)^{\frac{1}{2}}]^{p_0}, \tag{7.28}$$

with

$$t = -p_0 \tag{7.29}$$

and

$$c_2 = 3(p_0^2 - 1), \tag{7.30a}$$
$$c_3 = p_0(p_0^2 - 1), \tag{7.30b}$$
$$c_4 = -\tfrac{3}{4}(p_0^2 - 1)^2. \tag{7.30c}$$

Similarly,

$$g'(\xi, \epsilon; p_0) \equiv \langle 0, 0; p_0, p_0; -p_0 - 1 \mid \xi, \epsilon; t, m = p_0\rangle$$
$$= c'(p_0)e^{-\xi}\xi^{-\frac{1}{2}}(\xi^2 - \epsilon^2)^{-\frac{1}{4}}[\xi - (\xi^2 - \epsilon^2)^{\frac{1}{2}}]^{p_0}, \tag{7.31}$$

with

$$t = p_0 \tag{7.32}$$

and

$$c_2 = 3(p_0^2 - 1), \tag{7.33a}$$
$$c_3 = -p_0(p_0^2 - 1), \tag{7.33b}$$
$$c_4 = -\tfrac{3}{4}(p_0^2 - 1)^2. \tag{7.33c}$$

We have discussed the $E^{\pm}$ series in great detail, since they are the easiest. In the E^+ series, λ is always positive, and "energy" is positive; in the E^- series, λ is always negative, and "energy" is negative. We observe that $p_0 = 0, \frac{1}{2}, 1, \frac{3}{2}, \cdots$, but, for $p_0 = 0$, $f(\xi, \epsilon; p_0) = g(\xi, \epsilon; p_0)$, and there is only one UIR in the E^+ series. (The same remark holds for the E^- series.)

The "higher states" of a UIR in the E^+ series can be obtained very easily (see Papers I and II),

$$|j = p_0 + \tfrac{1}{2}, \mu = p_0 + \tfrac{1}{2}; k = \tfrac{1}{2}, \nu = -\tfrac{1}{2}; \lambda = p_0 + 2\rangle$$
$$= -i(2p_0 + 1)^{-\frac{1}{2}}P_+ |j = p_0, \mu = p_0; k = 0, \nu = 0; \lambda = p_0 + 1\rangle, \tag{7.34}$$

and, with the repeated application of P_+, Q_+, S_+, and T_+ together with Appendix A, we can calculate all matrix elements of the form $\langle j, \mu; k, \nu; \lambda \mid \xi, \epsilon; t, m\rangle$. The results are expressible in terms of derivatives of Eq. (7.20).

B. The "Timelike" Representations of $E(3, 1)$

We study next the more complicated "timelike" representations of $E(3, 1)$ which were treated in Sec. 4. Let us consider the discrete series $D^{\pm}$ of $SU(2, 2)$. We have no *a priori* knowledge that we should make the reduction with respect to "timelike" representations of $E(3, 1)$ rather than "spacelike" representations; only the result will tell that, indeed, "timelike" representations are the relevant choice.

1. *Discrete Series (the D^+ Series), Class IV*

We now discuss the discrete series (the D^+ series), class IV (Paper II) (see Theorem 4, Paper I). The three Casimir operators are [see Eqs. (6.8)–(6.10) of Paper I]

$$c_2 = 2j_m(j_m + 1) + 2k_m(k_m + 1) + \lambda_m(\lambda_m + 4), \tag{4.33'}$$

$$c_3 = -(\lambda_m + 2)(j_m - k_m)(j_m + k_m + 1), \tag{4.34'}$$

$$c_4 = \tfrac{1}{4}[(\lambda_m + 2)^2 - 4j_m(j_m + 1)] \times [(\lambda_m + 2)^2 - 4k_m(k_m + 1)] - (\lambda_m + 2)^2, \tag{4.35'}$$

where $j_m, k_m = 0, \frac{1}{2}, 1, \frac{3}{2}, \cdots$ and $\lambda_m = j_m + k_m + s_m$, $s_m = 0, 1, 2, \cdots$. (See Theorem 3 of Paper I,

where we see that $s_m = -1$ is also allowed; furthermore, if either j_m or k_m equals zero, $s_m = -2$ is also allowed.) Consider one of the corner states (states of highest weight) of a UIR of D^+ [see Eq. (I.6.11)], $|j = j_m,\ \mu = j_m;\ k = k_m,\ \nu = k_m;\ \lambda = \lambda_m + 4\rangle$. This state has the special property that

$$\begin{aligned} P_- |j_m, j_m; k_m, k_m; \lambda_m + 4\rangle &= Q_- |j_m, j_m; k_m, k_m; \lambda_m + 4\rangle \\ &= S_- |j_m, j_m; k_m, k_m; \lambda_m + 4\rangle \\ &= T_- |j_m, j_m; k_m, k_m; \lambda_m + 4\rangle = 0. \end{aligned} \tag{7.35}$$

We again have [analogous to Eqs. (7.7)–(7.10)]

$$\langle j_m, j_m; k_m, k_m; \lambda_m + 4| \mathfrak{L}_- |\eta, \xi, \epsilon; s, t, m\rangle = 0, \tag{7.36}$$

$$\begin{aligned} &f(\eta, \xi, \epsilon; s, t) \\ &\equiv \langle j_m, j_m; k_m, k_m; \lambda_m + 4 | \eta, \xi, \epsilon; s, t, m = j_m + k_m\rangle, \end{aligned} \tag{7.37}$$

and

$$\begin{aligned} f(\eta, \xi, \epsilon; s, t) = C(\eta, \xi; s, t)\epsilon^{j_m + k_m - t}(\xi^2 - \epsilon^2)^{-\frac{1}{4}} \\ \times [\xi + (\xi^2 - \epsilon^2)^{\frac{1}{2}}]^t. \end{aligned} \tag{7.38}$$

Appendix A now provides us with eight useful relations,

$$\tfrac{1}{2}(\mathfrak{T}_+ - \mathfrak{R}_+) = -i(J_+ + K_+), \tag{7.39}$$

$$\tfrac{1}{2}(\mathfrak{T}_- - \mathfrak{R}_-) = i(J_- + K_-), \tag{7.40}$$

$$\tfrac{1}{2}(\mathfrak{T}_5 - \mathfrak{R}_5) = J_3 - K_3, \tag{7.41}$$

$$\tfrac{1}{2}(\mathfrak{T}_6 + \mathfrak{R}_6) = R_0, \tag{7.42}$$

$$\mathfrak{L}_{+6} - \tfrac{1}{2}i(\mathfrak{T}_+ + \mathfrak{R}_+) = 2S_+, \tag{7.43}$$

$$\mathfrak{L}_{-6} - \tfrac{1}{2}i(\mathfrak{T}_- + \mathfrak{R}_-) = 2T_+, \tag{7.44}$$

$$i\mathfrak{L}_{56} + \tfrac{1}{2}(\mathfrak{T}_5 + \mathfrak{R}_5) = P_+ + Q_+, \tag{7.45}$$

$$-iL_{03} + \tfrac{1}{2}(\mathfrak{T}_6 - \mathfrak{R}_6) = P_+ - Q_+. \tag{7.46}$$

For unitary irreducible representations, Eqs. (7.35), together with Eq. (7.45), give us

$$\begin{aligned} &\langle j_m, j_m; k_m, k_m; \lambda_m + 4| i\mathfrak{L}_{56} + \tfrac{1}{2}(\mathfrak{T}_5 + \mathfrak{R}_5) \\ &\quad \times |\eta, \xi, \epsilon; s, t, m = j_m + k_m\rangle = 0, \end{aligned} \tag{7.47a}$$

which upon using Eq. (7.41) becomes

$$\begin{aligned} &\langle j_m, j_m; k_m, k_m; \lambda_m + 4| i\mathfrak{L}_{56} + \mathfrak{T}_5 - (j_m - k_m) \\ &\quad \times |\eta, \xi, \epsilon; s, t, m = j_m + k_m\rangle = 0. \end{aligned} \tag{7.47b}$$

Equation (3.11) now takes Eq. (7.47b) into

$$\begin{aligned} &\Big((\xi^2 + \eta^2)^{\frac{1}{2}}(\xi^2 - \epsilon^2)^{\frac{1}{2}}\frac{1}{\xi}\frac{\partial}{\partial \xi} \\ &\quad + \tfrac{1}{2}(\xi^2 - \epsilon^2)^{\frac{1}{2}}(\xi^2 + \eta^2)^{-\frac{1}{2}} + \tfrac{1}{2}(\xi^2 + \eta^2)^{\frac{1}{2}}(\xi^2 - \epsilon^2)^{-\frac{1}{2}} \\ &\quad + (\xi^2 - \epsilon^2)^{\frac{1}{2}} - (j_m - k_m)\Big) f(\eta, \xi, \epsilon; s, t) \\ &\quad - \tfrac{1}{2}[(s - t)(s + t + 1)]^{\frac{1}{2}} \\ &\quad \times \epsilon\eta\xi^{-2} f(\eta, \xi, \epsilon; s, t + 1) \\ &\quad + \tfrac{1}{2}[(s + t)(s - t + 1)]^{\frac{1}{2}} \\ &\quad \times \epsilon\eta\xi^{-2} f(\eta, \xi, \epsilon; s, t - 1) = 0, \end{aligned} \tag{7.47c}$$

which upon using Eq. (7.38) becomes

$$\begin{aligned} &\Big[(\xi^2 - \epsilon^2)^{\frac{1}{2}}\Big((\xi^2 + \eta^2)^{\frac{1}{2}}\frac{1}{\xi}\frac{\partial}{\partial \xi} + \frac{(\xi^2 + \eta^2)^{-\frac{1}{2}}}{2} + 1\Big) \\ &\quad + t\xi^{-1}(\xi^2 + \eta^2)^{\frac{1}{2}} - (j_m - k_m)\Big] C(\eta, \xi; s, t) \\ &\quad - \tfrac{1}{2}[(s - t)(s + t + 1)]^{\frac{1}{2}}\eta\xi^{-2}[\xi + (\xi^2 - \epsilon^2)^{\frac{1}{2}}] \\ &\quad \times C(\eta, \xi; s, t + 1) \\ &\quad + \tfrac{1}{2}[(s + t)(s - t + 1)]^{\frac{1}{2}}\eta\xi^{-2}[\xi - (\xi^2 - \epsilon^2)^{\frac{1}{2}}] \\ &\quad \times C(\eta, \xi; s, t - 1) = 0. \end{aligned} \tag{7.47d}$$

Since $C(\eta, \xi; s, t)$ is independent of ϵ, Eq. (7.47d) breaks up into two equations,

$$\begin{aligned} &\Big((\xi^2 + \eta^2)^{\frac{1}{2}}\frac{1}{\xi}\frac{\partial}{\partial \xi} + \frac{(\xi^2 + \eta^2)^{-\frac{1}{2}}}{2} + 1\Big) C(\eta, \xi; s, t) \\ &\quad - \tfrac{1}{2}[(s - t)(s + t + 1)]^{\frac{1}{2}}\eta\xi^{-2} C(\eta, \xi; s, t + 1) \\ &\quad - \tfrac{1}{2}[(s + t)(s - t + 1)]^{\frac{1}{2}}\eta\xi^{-2} C(\eta, \xi; s, t - 1) = 0 \end{aligned} \tag{7.48}$$

and

$$\begin{aligned} &[t\xi^{-1}(\xi^2 + \eta^2)^{\frac{1}{2}} - (j_m - k_m)] C(\eta, \xi; s, t) \\ &\quad - \tfrac{1}{2}[(s - t)(s + t + 1)]^{\frac{1}{2}}\eta\xi^{-1} C(\eta, \xi; s, t + 1) \\ &\quad + \tfrac{1}{2}[(s + t)(s - t + 1)]^{\frac{1}{2}}\eta\xi^{-1} C(\eta, \xi; s, t - 1) = 0. \end{aligned} \tag{7.49}$$

Let

$$\begin{aligned} &C(\eta, \xi; s, t) \\ &\quad = \exp[-(\xi^2 + \eta^2)^{\frac{1}{2}}](\xi^2 + \eta^2)^{-\frac{1}{4}} h(\eta, \xi; s, t). \end{aligned} \tag{7.50}$$

The sum and difference of Eqs. (7.48) and (7.49) become

$$\begin{aligned} &\Big((\xi^2 + \eta^2)^{\frac{1}{2}}\frac{\partial}{\partial \xi} + t\xi^{-1}(\xi^2 + \eta^2)^{\frac{1}{2}} - (j_m - k_m)\Big) \\ &\quad \times h(\eta, \xi; s, t) \\ &\quad = [(s - t)(s + t + 1)]^{\frac{1}{2}}\eta\xi^{-1} h(\eta, \xi; s, t + 1) \end{aligned} \tag{7.51}$$

and

$$\left((\xi^2+\eta^2)^{\frac{1}{2}}\frac{\partial}{\partial\xi} - t\xi^{-1}(\xi^2+\eta^2)^{\frac{1}{2}} + (j_m - k_m)\right) \times h(\eta, \xi; s, t) = [(s+t)(s-t+1)]^{\frac{1}{2}}\eta\xi^{-1}h(\eta, \xi; s, t-1). \quad (7.52)$$

These two coupled first-order partial differential equations can now be reduced to a second-order differential equation:

$$\left((\xi^2+\eta^2)\frac{\partial^2}{\partial\xi^2} + \xi\frac{\partial}{\partial\xi} + 2(j_m - k_m)t\xi^{-1}(\xi^2+\eta^2)^{\frac{1}{2}} - (j_m - k_m)^2 - t^2 - s(s+1)\eta^2\xi^{-2}\right) \times h(\eta, \xi; s, t) = 0. \quad (7.53)$$

Next, we consider Eqs. (7.46) and (7.42):

$$\langle j_m, j_m; k_m, k_m; \lambda_m + 4 | \tfrac{1}{2}(\mathfrak{I}_6 - \mathfrak{R}_6) - iL_{03} \times |\eta, \xi, \epsilon; s, t, m = j_m + k_m\rangle = 0, \quad (7.54)$$

$$\langle j_m, j_m; k_m, k_m; \lambda_m + 4 | \mathfrak{I}_6 - (\lambda_m + 4) - iL_{03} \times |\eta, \xi, \epsilon; s, t, m = j_m + k_m\rangle = 0, \quad (7.55)$$

which upon using Eq. (4.9) becomes

$$\left(\eta\frac{\partial}{\partial\eta} + \xi\frac{\partial}{\partial\xi} + \epsilon\frac{\partial}{\partial\epsilon} + 3 + (\xi^2+\eta^2)^{\frac{1}{2}} - (\lambda_m + 4)\right) \times f(\eta, \xi, \epsilon; s, t) = 0 \quad (7.56)$$

or

$$\left(\eta\frac{\partial}{\partial\eta} + \xi\frac{\partial}{\partial\xi} + (j_m + k_m) - (\lambda_m + 2)\right) \times h(\eta, \xi; s, t) = 0. \quad (7.57)$$

Before solving for Eqs. (7.53) and (7.57), we remark that Eqs. (7.39), (7.40), (7.43), and (7.44) do not give us any new information; they again provide us with Eq. (7.53). The consideration of the Casimir operators now becomes important. We need to check only with C_2:

$$\langle j_m, j_m; k_m, k_m; \lambda_m + 4 | C_2 | \eta, \xi, \epsilon; s, t, m = j_m + k_m\rangle = c_2 f(\eta, \xi, \epsilon; s, t). \quad (7.58)$$

Using Eqs. (4.19), (3.3)–(3.11), and (4.9) together with Eqs. (7.39)–(7.46), we get, after some computation,

$$\left((\xi^2+\eta^2)\frac{\partial^2}{\partial\xi^2} + \xi\frac{\partial}{\partial\xi} + 2(j_m - k_m)t\xi^{-1}(\xi^2+\eta^2)^{\frac{1}{2}} - t^2 - s(s+1)\eta^2\xi^{-2} - c_2 + 2(j_m + k_m) + (j_m + k_m)^2 + \lambda_m(\lambda_m + 4)\right)h(\eta, \xi; s, t) = 0, \quad (7.59)$$

which upon comparison with Eq. (7.53) gives

$$c_2 = 2j_m(j_m + 1) + 2k_m(k_m + 1) + \lambda_m(\lambda_m + 4), \quad (7.60)$$

which is the result we have in Eq. (4.33′).

Let

$$h(\eta, \xi; s, t) = \xi^{-s}[(\xi^2+\eta^2)^{\frac{1}{2}} + \xi]^{j_m - k_m}g(\eta, \xi; s, t). \quad (7.61)$$

Equation (7.51) now becomes

$$\left((\xi^2+\eta^2)^{\frac{1}{2}}\frac{\partial}{\partial\xi} - (s-t)\xi^{-1}(\xi^2+\eta^2)^{\frac{1}{2}}\right)g(\eta, \xi; s, t) = [(s-t)(s+t+1)]^{\frac{1}{2}}\eta\xi^{-1}g(\eta, \xi; s, t+1), \quad (7.62)$$

which is solved immediately for $t = s$:

$$g(\eta, \xi; s, t = s) = G(\eta; s). \quad (7.63)$$

Substituting Eqs. (7.61) and (7.63) into (7.57), we get

$$\left(\eta\frac{\partial}{\partial\eta} + 2j_m - s - (\lambda_m + 2)\right)G(\eta; s) = 0 \quad (7.64)$$

and

$$G(\eta, s) = c\eta^{s - 2j_m + (\lambda_m + 2)}. \quad (7.65)$$

In Eq. (7.37), $m = j_m + k_m$; therefore, $s \geq j_m + k_m$. From Eqs. (4.28) and (4.38) we see that

$$\gamma(s = j_m + k_m) = 0;$$

hence, $s = j_m + k_m$, and no other value of s is allowed. We have, finally,

$$f(\eta, \xi, \epsilon; s = j_m + k_m, t = s) = c\exp[-(\xi^2+\eta^2)^{\frac{1}{2}}](\xi^2+\eta^2)^{-\frac{1}{4}}\eta^{(\lambda_m+2)-(j_m-k_m)} \times \xi^{-(j_m+k_m)}[(\xi^2+\eta^2)^{\frac{1}{2}} + \xi]^{j_m-k_m}(\xi^2 - \epsilon^2)^{-\frac{1}{4}} \times [\xi + (\xi^2 - \epsilon^2)^{\frac{1}{2}}]^{j_m+k_m}. \quad (7.66)$$

The general case for $f(\eta, \xi, \epsilon; s = j_m + k_m, t)$ with $-s \leq t \leq s$ can be obtained by the repeated application of Eq. (7.52); the results, however, are omitted here.

Next, we remark on the general matrix element

$$\langle j, \mu; k, \nu; \lambda | \eta, \xi, \epsilon; s, t, m = \mu + \nu\rangle.$$

For the nondegenerate unitary irreducible representations in the D^+ series, the general matrix element is difficult to compute due to the presence of the additional label α in $|j, \mu; k, \nu; \lambda; \alpha\rangle$ (see Paper I). However, for degenerate UIR's, the general matrix element is easy to calculate with the help of Eqs. (II.1.7) and (II.1.8). (We remind the reader that degenerate UIR's in the D^+ series have either j_m or k_m equal to zero.)

2. *UIR's in the D^- Series*

The treatment of the UIR's in the D^- series is similar to that given above for the D^+ series. The only difference is that the "energy" is now negative [compare with Eq. (7.27)]:

$$\mathfrak{F}_6 |\eta, \xi, \epsilon; s, t, m\rangle = -(\xi^2 + \eta^2)^{\frac{1}{2}} |\eta, \xi, \epsilon; s, t, m\rangle. \tag{7.67}$$

Finally, we must answer the important question of why we have made the reduction of the discrete series $D^{\pm}$ of $SU(2,2)$ with respect to the "timelike" representations of $E(3,1)$. Indeed, instead of Eqs. (7.37)–(7.59), we could write

$$\tilde{f}(\omega, \xi, \epsilon; r, t) \equiv \langle j_m, j_m; k_m, k_m; \lambda_m + 4 \,|\, \omega, \xi, \epsilon; r, t, m = j_m + k_m\rangle, \tag{7.37'}$$

$$\tilde{f}(\omega, \xi, \epsilon; r, t) = \tilde{c}(\omega, \xi; r, t)\epsilon^{j_m+k_m-t}(\xi^2 - \epsilon^2)^{-\frac{1}{4}}[\xi + (\xi^2 - \epsilon^2)^{\frac{1}{2}}]^t, \tag{7.38'}$$

$$\tilde{c}(\omega, \xi, \epsilon; r, t) = \exp[-(\xi^2 - \omega^2)^{\frac{1}{2}}](\xi^2 - \omega^2)^{-\frac{1}{4}}\tilde{h}(\omega, \xi; r, t), \tag{7.50'}$$

$$\left((\xi^2 - \omega^2)^{\frac{1}{2}}\frac{\partial}{\partial \xi} + t\xi^{-1}(\xi^2 - \omega^2)^{\frac{1}{2}} - (j_m - k_m)\right) \times \tilde{h}(\omega, \xi; r, t) = -[(t + r)(t - r + 1)]^{\frac{1}{2}}\omega\xi^{-1}\tilde{h}(\omega, \xi; r, t + 1), \tag{7.51'}$$

$$\left((\xi^2 - \omega^2)^{\frac{1}{2}}\frac{\partial}{\partial \xi} - t\xi^{-1}(\xi^2 - \omega^2)^{\frac{1}{2}} + (j_m - k_m)\right) \times \tilde{h}(\omega, \xi; r, t) = -[(t - r)(t + r - 1)]^{\frac{1}{2}}\omega\xi^{-1}\tilde{h}(\omega, \xi; r, t - 1), \tag{7.52'}$$

and

$$\left(\omega\frac{\partial}{\partial \omega} + \xi\frac{\partial}{\partial \xi} + (j_m + k_m) - (\lambda_m + 2)\right) \times \tilde{h}(\omega, \xi; r, t) = 0, \tag{7.57'}$$

which are the results obtained by reducing the D^+ series with respect to the "spacelike" representations of $E(3,1)$. Let

$$\tilde{h}(\omega, \xi; r, t) = \xi^r[\xi + (\xi^2 - \omega^2)^{\frac{1}{2}}]^{-(j_m-k_m)}\tilde{g}(\omega, \xi; r, t). \tag{7.61'}$$

Equation (7.52') becomes

$$\left((\xi^2 - \omega^2)^{\frac{1}{2}}\frac{\partial}{\partial \xi} + (r - t)\xi^{-1}(\xi^2 - \omega^2)^{\frac{1}{2}}\right)\tilde{g}(\omega, \xi; r, t) = -[(t - r)(t + r - 1)]^{\frac{1}{2}}\omega\xi^{-1}\tilde{g}(\omega, \xi; r, t - 1), \tag{7.62'}$$

which is solved immediately for $t = r$:

$$\hat{g}(\omega, \xi; r, t = r) = \tilde{G}(\omega; r). \tag{7.63'}$$

Substituting Eqs. (7.61') and (7.63') into (7.57'), we get

$$\left(\omega\frac{\partial}{\partial \omega} + 2k_m + r - (\lambda_m + 2)\right)\tilde{G}(\omega; r) = 0 \tag{7.64'}$$

and

$$\tilde{G}(\omega; r) = \tilde{c}(r)\omega^{-r-2k_m+(\lambda_m+2)} \tag{7.65'}$$

We have, finally,

$$\tilde{f}(\omega, \xi, \epsilon; r, t = r) = \tilde{c}(r)\exp[-(\xi^2 - \omega^2)^{\frac{1}{2}}](\xi^2 - \omega^2)^{-\frac{1}{4}}\omega^{-r-2k_m+(\lambda_m+2)}\xi^r \times [\xi + (\xi^2 - \omega^2)^{\frac{1}{2}}]^{-(j_m-k_m)}(\xi^2 - \epsilon^2)^{-\frac{1}{4}} \times [\xi + (\xi^2 - \epsilon^2)^{\frac{1}{2}}]^{j_m+k_m}. \tag{7.66'}$$

So far, the derivation for Eq. (7.66') parallels that for Eq. (7.66), and both have a claim to our attention. The important difference comes about when we consider the problem of convergence. Since

$$\langle j_m, j_m; k_m, k_m; \lambda_m + 4 \,|\, j_m, j_m; k_m, k_m; \lambda_m + 4\rangle$$

is normalized, the integral

$$\int_0^\infty d\eta^2 \int_0^\infty d\xi^2 \int_0^{\xi^2} d\epsilon^2 \, |f(\eta, \xi, \epsilon; s, t)|^2$$

must be finite, which requires [from Eq. (7.66)] that

$$(\lambda_m + 2) - (j_m - k_m) + 1 > 0, \tag{7.68}$$

a condition satisfied automatically since $\lambda_m = j_m + k_m + s_m$. On the other hand, the integral

$$\int_0^\infty d\xi^2 \int_0^{\xi^2} d\omega^2 \int_0^{\xi^2} d\epsilon^2 \, |\tilde{f}(\omega, \xi, \epsilon; r, t)|^2$$

is infinite since $\tilde{f}(\omega, \xi, \epsilon; r, t)$ is unbounded as $\omega \to \xi$, $\xi \to \infty$. Therefore, the UIR's in the $D^{\pm}$ series are not reducible with respect to the "spacelike" representation of $E(3,1)$, and

$$\tilde{c}(r) = 0. \tag{7.69}$$

3. *The Most Degenerate Principal Continuous Series*

We next come to the study of the most degenerate principal continuous series, class I (Paper II). The three Casimir operators are

$$c_2 = -4 - \rho^2, \tag{7.70a}$$

$$c_3 = 0, \tag{7.70b}$$

$$c_4 = \tfrac{1}{4}\rho^4 + \rho^2, \tag{7.70c}$$

where $\rho > 0$. The allowed state of $SU(2) \times SU(2) \times U(1)$ obeys the restrictions of either (a) $p + \lambda = \pm$ even integer or (b) $p + \lambda = \pm$ odd integer, with $p = 0, 1, 2, \cdots$ [see Eqs. (II.3.45) and (II.3.46)].

Let us start with one of the "ground" states $|j = 0, \mu = 0; k = 0, \nu = 0; \lambda\rangle$, and define

$$f(\eta, \xi, \epsilon; \lambda) \equiv \langle j = 0, \mu = 0; k = 0, \nu = 0; \lambda \,|\, \eta, \xi, \epsilon; s = 0, t = 0, m = 0\rangle. \quad (7.71)$$

From Eqs. (4.28)–(4.31), we see that $A = i\rho$, $B = 0$, $C = 1$, and $\gamma^2(s) = 0$ for $s = 0$. Therefore, $s = 0$ always for the series under consideration.

From

$$\tfrac{1}{2}(\mathfrak{T}_6 + \mathfrak{R}_6) = R_0 \quad (7.42)$$

and Eq. (4.14), we get

$$\Big[(\xi^2+\eta^2)^{\frac{1}{2}}\Big(\frac{\partial^2}{\partial\eta^2}+\frac{\partial^2}{\partial\xi^2}+\frac{\partial^2}{\partial\epsilon^2}+2\frac{\epsilon}{\xi}\frac{\partial}{\partial\xi}\frac{\partial}{\partial\epsilon}\Big) + [\eta(\xi^2+\eta^2)^{-\frac{1}{2}}+\eta^{-1}(\xi^2+\eta^2)^{\frac{1}{2}}]\frac{\partial}{\partial\eta} + \xi^{-1}[4(\xi^2+\eta^2)^{\frac{1}{2}}-\eta^2(\xi^2+\eta^2)^{-\frac{1}{2}}]\frac{\partial}{\partial\xi} + [\epsilon(\xi^2+\eta^2)^{-\frac{1}{2}}+\epsilon^{-1}(\xi^2+\eta^2)^{\frac{1}{2}}]\frac{\partial}{\partial\epsilon} - \tfrac{1}{4}(\xi^2+\eta^2)^{\frac{1}{2}}(\xi^2-\epsilon^2)^{-1}+\tfrac{9}{4}(\xi^2+\eta^2)^{-\frac{1}{2}} - (\xi^2+\eta^2)^{\frac{1}{2}}+2\lambda+\rho^2\eta^{-2}(\xi^2+\eta^2)^{\frac{1}{2}}\Big] \times f(\eta, \xi, \epsilon; \lambda) = 0. \quad (7.72)$$

From

$$\tfrac{1}{2}(\mathfrak{T}_5 - \mathfrak{R}_5) = J_3 - K_3 \quad (7.41)$$

and Eq. (4.11), we get

$$\Big[(\xi^2-\epsilon^2)^{\frac{1}{2}}\Big(-\frac{\partial^2}{\partial\eta^2}+\frac{\partial^2}{\partial\xi^2}-\frac{\partial^2}{\partial\epsilon^2}+2\frac{\eta}{\xi}\frac{\partial}{\partial\eta}\frac{\partial}{\partial\xi}\Big) + [\eta(\xi^2-\epsilon^2)^{-\frac{1}{2}}-\eta^{-1}(\xi^2-\epsilon^2)^{\frac{1}{2}}]\frac{\partial}{\partial\eta} + \xi^{-1}[4(\xi^2-\epsilon^2)^{\frac{1}{2}}+\epsilon^2(\xi^2-\epsilon^2)^{-\frac{1}{2}}]\frac{\partial}{\partial\xi} + [\epsilon(\xi^2-\epsilon^2)^{-\frac{1}{2}}-\epsilon^{-1}(\xi^2-\epsilon^2)^{\frac{1}{2}}]\frac{\partial}{\partial\epsilon} - \tfrac{1}{4}(\xi^2-\epsilon^2)^{\frac{1}{2}}(\xi^2+\eta^2)^{-1}+\tfrac{9}{4}(\xi^2-\epsilon^2)^{-\frac{1}{2}} - \rho^2\eta^{-2}(\xi^2-\epsilon^2)^{\frac{1}{2}}-(\xi^2-\epsilon^2)^{\frac{1}{2}}\Big] f(\eta, \xi, \epsilon; \lambda) = 0. \quad (7.73)$$

Substituting

$$f(\eta, \xi, \epsilon; \lambda) = (\xi^2 - \epsilon^2)^{-\frac{1}{4}} C(\eta, \xi; \lambda) \quad (7.74)$$

into Eqs. (7.72) and (7.73), we get

$$\Big[(\xi^2+\eta^2)^{\frac{1}{2}}\Big(\frac{\partial^2}{\partial\eta^2}+\frac{\partial^2}{\partial\xi^2}\Big) + [\eta(\xi^2+\eta^2)^{-\frac{1}{2}}+\eta^{-1}(\xi^2+\eta^2)^{\frac{1}{2}}]\frac{\partial}{\partial\eta} + \xi^{-1}[3(\xi^2+\eta^2)^{\frac{1}{2}}-\eta^2(\xi^2+\eta^2)^{-\frac{1}{2}}]\frac{\partial}{\partial\xi} + \tfrac{7}{4}(\xi^2+\eta^2)^{-\frac{1}{2}} + \rho^2\eta^{-2}(\xi^2+\eta^2)^{\frac{1}{2}}+2\lambda-(\xi^2+\eta^2)^{\frac{1}{2}}\Big] C(\eta, \xi; \lambda) = 0 \quad (7.75)$$

and

$$\Big(-\frac{\partial^2}{\partial\eta^2}+\frac{\partial^2}{\partial\xi^2}-\frac{1}{\eta}\frac{\partial}{\partial\eta}+\frac{3}{\xi}\frac{\partial}{\partial\xi}+2\frac{\eta}{\xi}\frac{\partial}{\partial\eta}\frac{\partial}{\partial\xi} - \frac{(\xi^2+\eta^2)^{-1}}{4}-\rho^2\frac{1}{\eta^2}-1\Big) C(\eta, \xi; \lambda) = 0. \quad (7.76)$$

The further substitution

$$C(\eta, \xi; \lambda) = (\xi^2 + \eta^2)^{-\frac{1}{4}} h(\eta, \xi; \lambda) \quad (7.77)$$

reduces Eqs. (7.75) and (7.76) to

$$\Big(\frac{\partial^2}{\partial\xi^2}+\frac{2}{\xi}\frac{\partial}{\partial\xi}-1+\frac{\eta}{\xi}\frac{\partial}{\partial\eta}\frac{\partial}{\partial\xi}+\lambda(\xi^2+\eta^2)^{-\frac{1}{2}}\Big) \times h(\eta, \xi; \lambda) = 0, \quad (7.78)$$

$$\Big(\frac{\partial^2}{\partial\eta^2}+\frac{1}{\eta}\frac{\partial}{\partial\eta}+\rho^2\frac{1}{\eta^2}-\frac{\eta}{\xi}\frac{\partial}{\partial\eta}\frac{\partial}{\partial\xi}+\lambda(\xi^2+\eta^2)^{-\frac{1}{2}}\Big) \times h(\eta, \xi; \lambda) = 0. \quad (7.79)$$

In order to make Eqs. (7.78) and (7.79) separable, we introduce the parabolic coordinates

$$x = (\xi^2 + \eta^2)^{\frac{1}{2}} + \xi, \quad (7.80a)$$

$$y = (\xi^2 + \eta^2)^{\frac{1}{2}} - \xi. \quad (7.80b)$$

Equations (7.78) and (7.79) now become

$$\Big[x\Big(\frac{\partial^2}{\partial x^2}+\frac{1}{x}\frac{\partial}{\partial x}-\frac{1}{4}+\frac{1}{4}\frac{\rho^2}{x^2}\Big) - y\Big(\frac{\partial^2}{\partial y^2}+\frac{1}{y}\frac{\partial}{\partial y}-\frac{1}{4}+\frac{1}{4}\frac{\rho^2}{y^2}\Big)\Big] h(x, y; \lambda) = 0, \quad (7.81)$$

$$\Big[x^2\Big(\frac{\partial^2}{\partial x^2}+\frac{2}{x}\frac{\partial}{\partial x}-\frac{1}{4}+\frac{1}{2}\frac{\lambda}{x}\Big) - y^2\Big(\frac{\partial^2}{\partial y^2}+\frac{2}{y}\frac{\partial}{\partial y}-\frac{1}{4}+\frac{1}{2}\frac{\lambda}{y}\Big)\Big] h(x, y; \lambda) = 0, \quad (7.82)$$

which are separable and antisymmetric in x and y. Since the differential equations are of the confluent hypergeometric kind in both variables, we try the ansatz[9]

$$h(x, y; \lambda) = \int_0^\infty dv\, e^{-\frac{1}{2}(x+y)\cosh v}(\coth \tfrac{1}{2}v)^\lambda g(u, v), \tag{7.83}$$

where $u = xy$. Substituting Eq. (7.83) into Eq. (7.82), we find that

$$g(u, v) = u^{-\frac{1}{2}}G(u^{\frac{1}{2}} \sinh v).$$

To determine $G(u^{\frac{1}{2}} \sinh v)$, we substitute Eqs. (7.83) into Eq. (7.81) and find

$$g(u, v) = \begin{cases} C_2 \sinh v K_{i\rho}(u^{\frac{1}{2}} \sinh v) \\ C_1 \sinh v I_{i\rho}(u^{\frac{1}{2}} \sinh v) \end{cases}, \tag{7.84}$$

where $I_\mu(z)$ and $K_\mu(z)$ are modified Bessel's functions of the first and second kind, respectively. The function $I_{i\rho}(u^{\frac{1}{2}} \sinh v)$ can be shown to be unacceptable since it leads to an $f(\eta, \xi, \epsilon; \lambda)$ which is not normalizable. The function $K_{i\rho}(u^{\frac{1}{2}} \sinh v)$ gives

$$h(x, y; \lambda) = C_2 \int_0^\infty dv\, e^{-\frac{1}{2}(x+y)\cosh v}(\coth \tfrac{1}{2}v)^\lambda \times \sinh v K_{i\rho}(x^{\frac{1}{2}}y^{\frac{1}{2}} \sinh v), \tag{7.85}$$

and is convergent for $\lambda \leq 0$. This integral can be performed, and we obtain

$$h(x, y; \lambda) = C(\lambda, \rho)[2/(x - y)][W_{\frac{1}{2}\lambda,\frac{1}{2}(i\rho+1)}(x)W_{\frac{1}{2}\lambda,\frac{1}{2}(i\rho-1)}(y) - W_{\frac{1}{2}\lambda,\frac{1}{2}(i\rho-1)}(x)W_{\frac{1}{2}\lambda,\frac{1}{2}(i\rho+1)}(y)], \tag{7.86}$$

where $W_{\kappa,\mu}(z)$ is Whittaker's function of the second kind. Finally, we have

$$f(\eta, \xi, \epsilon; \lambda) = C(\lambda, \rho)(\xi^2 - \epsilon^2)^{-\frac{1}{4}}(\xi^2 + \eta^2)^{-\frac{1}{4}} \times \xi^{-1}[W_{\frac{1}{2}\lambda,\frac{1}{2}(i\rho+1)}(x)W_{\frac{1}{2}\lambda,\frac{1}{2}(i\rho-1)}(y) - W_{\frac{1}{2}\lambda,\frac{1}{2}(i\rho-1)}(x)W_{\frac{1}{2}\lambda,\frac{1}{2}(i\rho+1)}(y)], \tag{7.87}$$

where $C(\lambda, \rho)$ is a normalization constant.[10]

4. *The Most Degenerate Complementary Continuous Series*

We next discuss the most degenerate complementary continuous series, class II (Paper II).

The three Casimir operators are

$$C_2 = -4 + \sigma^2, \tag{7.88a}$$

$$C_3 = 0, \tag{7.88b}$$

$$C_4 = \tfrac{1}{4}\sigma^4 - \sigma^2, \tag{7.88c}$$

where $0 \leq \sigma < 1$. The allowed states of $SU(2) \times SU(2) \times U(1)$ obey the restriction $p + \lambda = \pm$ odd integer, with $p = 0, 1, 2, \cdots$ [see Eq. (II.3.51)].

Equations (7.88) are similar to Eqs. (7.70) and can be regarded as analytic continuations of the latter. Substituting $i\rho \to \sigma$, we get immediately

$$f(\eta, \xi, \epsilon; \lambda) = C(\lambda, \sigma)(\xi^2 - \epsilon^2)^{-\frac{1}{4}}(\xi^2 + \eta^2)^{-\frac{1}{4}} \times \xi^{-1}[W_{\frac{1}{2}\lambda,\frac{1}{2}(1+\sigma)}(x)W_{\frac{1}{2}\lambda,\frac{1}{2}(1-\sigma)}(y) - W_{\frac{1}{2}\lambda,\frac{1}{2}(1-\sigma)}(x)W_{\frac{1}{2}\lambda,\frac{1}{2}(1+\sigma)}(y)]. \tag{7.89}$$

However, in contrast to Eq. (7.87), where λ can be either even or odd integers, here in Eq. (7.89) λ is restricted to odd integers.[10]

C. The "Spacelike" Representations of $E(3, 1)$

We study next the "spacelike" representations of $E(3, 1)$ which were treated in Sec. 5. There are five cases to be studied.

1. *Principal Series, Class XIII*

We start with the principal series, class XIII (Paper II). The three Casimir operators are

$$c_2 = p_0^2 - 2\rho^2 - \tfrac{9}{2}, \tag{7.90a}$$

$$c_3 = \pm p_0(\rho^2 + \tfrac{1}{4}), \tag{7.90b}$$

$$c_4 = \tfrac{1}{4}p_0^4 + p_0^2(\rho^2 - \tfrac{3}{4}), \tag{7.90c}$$

where $p_0 = 0, \frac{1}{2}, 1, \frac{3}{2}, \cdots$, $\rho > 0$. Comparing with Eqs. (4.29)–(4.31), we see that

$$A = \mp p_0, \tag{7.91a}$$

$$B = -\tfrac{1}{2} \pm i\rho, \tag{7.91b}$$

$$C = B + 1 = \tfrac{1}{2} \pm i\rho. \tag{7.91c}$$

The allowed states of the maximal compact subgroup $SU(2) \times SU(2) \times U(1)$ obey the restriction $p - \lambda \geq 0$ and $p + \lambda \geq 0$, where $p = j + k$ [see Eqs. (II.3.14) and (II.3.15)].

Let us specialize to $c_3 = p_0(\rho^2 + \frac{1}{4})$, $A = -p_0$, and consider the corner state $|j = p_0, \mu = p_0; k = 0, \nu = 0; \lambda = p_0\rangle$ [see Eq. (II.3.15)]. From

$$\tfrac{1}{2}(\mathfrak{T}_5 - \mathfrak{R}_5) = J_3 - K_3, \tag{7.41}$$

we have

$$\langle j = p_0, \mu = p_0; k = 0, \nu = 0; \lambda = p_0| \times \mathfrak{R}_5 - \mathfrak{T}_5 + 2p_0 |\omega, \xi, \epsilon; r, t, m = p_0\rangle = 0. \tag{7.92}$$

Next, we consider Eq. (4.28) with the appropriate change $s + 1 \to r$:

$$(2r - 1)(2r + 1)r^2\gamma^2(r) = (r^2 - A^2)(r^2 - B^2)(r^2 - C^2), \tag{7.93}$$

where A, B, and C are given by Eqs. (7.91). In Sec. 5 we learned from Bargmann that for $r = \frac{1}{2} \pm i\rho$ (the

odd continuous series $C_q^{\frac{1}{2}}$ and *part* of the even continuous series C_q^0) both $\gamma(r)$ and $\gamma(r-1)$ are automatically zero, and r is fixed for the principal series.

Let

$$\langle j = p_0, \mu = p_0; k = 0, \nu = 0; \lambda = p_0 \,|\, \omega, \xi, \epsilon; r = \tfrac{1}{2} \pm i\rho, t, m = p_0 \rangle \equiv f(\omega, \xi, \epsilon; r, t). \tag{7.94}$$

Equation (7.92) with the help of Eq. (4.11) (we change $\eta \to i\omega$, $s + 1 \to r$) can now be written as

$$\begin{aligned}
&\left[(\xi^2-\epsilon^2)^{\frac{1}{2}}\left(\frac{\partial^2}{\partial\omega^2}+\frac{\partial^2}{\partial\xi^2}-\frac{\partial^2}{\partial\epsilon^2}+2\frac{\omega}{\xi}\frac{\partial}{\partial\omega}\frac{\partial}{\partial\xi}\right)+[\omega(\xi^2-\epsilon^2)^{-\frac{1}{2}}+\omega^{-1}(\xi^2-\epsilon^2)^{\frac{1}{2}}]\frac{\partial}{\partial\omega}\right.\\
&\quad+\xi^{-1}[4(\xi^2-\epsilon^2)^{\frac{1}{2}}+\epsilon^2(\xi^2-\epsilon^2)^{-\frac{1}{2}}]\frac{\partial}{\partial\xi}+[\epsilon(\xi^2-\epsilon^2)^{-\frac{1}{2}}-\epsilon^{-1}(\xi^2-\epsilon^2)^{\frac{1}{2}}]\frac{\partial}{\partial\epsilon}\\
&\quad+p_0^2\epsilon^{-2}(\xi^2-\epsilon^2)^{\frac{1}{2}}-2p_0t\xi\epsilon^{-2}+t^2(\xi^2-\epsilon^2)^{\frac{1}{2}}(-\omega^{-2}+2\xi^{-2}+\epsilon^{-2})\\
&\quad+(\rho^2+\tfrac{1}{4})\xi^{-2}(\xi^2-\epsilon^2)^{\frac{1}{2}}+2p_0t\omega^{-2}\xi^{-1}(\xi^2-\epsilon^2)^{\frac{1}{2}}(\xi^2-\omega^2)^{\frac{1}{2}}-\tfrac{1}{4}(\xi^2-\epsilon^2)^{\frac{1}{2}}(\xi^2-\omega^2)^{-1}+\tfrac{9}{4}(\xi^2-\epsilon^2)^{-\frac{1}{2}}\\
&\quad\left.-p_0^2\omega^{-2}(\xi^2-\epsilon^2)^{\frac{1}{2}}-(\xi^2-\epsilon^2)^{\frac{1}{2}}+2p_0\right]f(\omega,\xi,\epsilon;r,t)+[(t+r)(t-r+1)]^{\frac{1}{2}}\epsilon\xi^{-1}\\
&\quad\times\left[-\xi^{-1}(\xi^2-\omega^2)^{\frac{1}{2}}\left(\frac{\partial}{\partial\omega}+\omega^{-1}(t+1)-\frac{\omega(\xi^2-\omega^2)^{-1}}{2}\right)+p_0\omega^{-1}\right]f(\omega,\xi,\epsilon;r,t+1)\\
&\quad+[(t-r)(t+r-1)]^{\frac{1}{2}}\epsilon\xi^{-1}\left[\xi^{-1}(\xi^2-\omega^2)^{\frac{1}{2}}\left(\frac{\partial}{\partial\omega}-\omega^{-1}(t-1)-\frac{\omega(\xi^2-\omega^2)^{-1}}{2}\right)+p_0\omega^{-1}\right]\\
&\quad\times f(\omega,\xi,\epsilon;r,t-1)=0.
\end{aligned}\tag{7.95}$$

Substituting

$$f(\omega,\xi,\epsilon;r,t)=\epsilon^{p_0-t}(\xi^2-\epsilon^2)^{-\frac{1}{4}}[\xi+(\xi^2-\epsilon^2)^{\frac{1}{2}}]^t C(\omega,\xi;r,t) \tag{7.96}$$

into Eq. (7.95), we get two independent coupled differential equations:

$$\begin{aligned}
&\left(\frac{\partial^2}{\partial\omega^2}+\frac{\partial^2}{\partial\xi^2}+2\frac{\omega}{\xi}\frac{\partial}{\partial\omega}\frac{\partial}{\partial\xi}+\frac{1}{\omega}\frac{\partial}{\partial\omega}+\frac{3}{\xi}\frac{\partial}{\partial\xi}+t^2(-\omega^{-2}+2\xi^{-2})+(\rho^2+\tfrac{1}{4})\xi^{-2}\right.\\
&\quad\left.-p_0^2\omega^{-2}-1+2p_0t\omega^{-2}\xi^{-1}(\xi^2-\omega^2)^{\frac{1}{2}}-\tfrac{1}{4}(\xi^2-\omega^2)^{-1}\right)C(\omega,\xi,r,t)\\
&\quad+[(t+r)(t-r+1)]^{\frac{1}{2}}\xi^{-1}\left[-\xi^{-1}(\xi^2-\omega^2)^{\frac{1}{2}}\left(\frac{\partial}{\partial\omega}+\omega^{-1}(t+1)-\frac{\omega(\xi^2-\omega^2)^{-1}}{2}\right)+p_0\omega^{-1}\right]C(\omega,\xi;r,t+1)\\
&\quad-[(t-r)(t+r-1)]^{\frac{1}{2}}\xi^{-1}\left[\xi^{-1}(\xi^2-\omega^2)^{\frac{1}{2}}\left(\frac{\partial}{\partial\omega}-\omega^{-1}(t-1)-\frac{\omega(\xi^2-\omega^2)^{-1}}{2}\right)+p_0\omega^{-1}\right]\\
&\quad\times C(\omega,\xi;r,t-1)=0,
\end{aligned}\tag{7.97}$$

$$\begin{aligned}
&\left[2t\left(\frac{\omega}{\xi}\frac{\partial}{\partial\omega}+\frac{\partial}{\partial\xi}\right)+3t\frac{1}{\xi}+2p_0\right]C(\omega,\xi;r,t)\\
&\quad+[(t+r)(t-r+1)]^{\frac{1}{2}}\left[-\xi^{-1}(\xi^2-\omega^2)^{\frac{1}{2}}\left(\frac{\partial}{\partial\omega}+\omega^{-1}(t+1)-\frac{\omega(\xi^2-\omega^2)^{-1}}{2}\right)+p_0\omega^{-1}\right]C(\omega,\xi;r,t+1)\\
&\quad+[(t-r)(t+r-1)]^{\frac{1}{2}}\left[\xi^{-1}(\xi^2-\omega^2)^{\frac{1}{2}}\left(\frac{\partial}{\partial\omega}-\omega^{-1}(t-1)-\frac{\omega(\xi^2-\omega^2)^{-1}}{2}\right)+p_0\omega^{-1}\right]C(\omega,\xi;r,t-1)=0.
\end{aligned}\tag{7.98}$$

Since $-p_0 \le t \le p_0$ in Eq. (7.94), we have

$$f(\omega,\xi,\epsilon;r,t=p_0+1)=f(\omega,\xi,\epsilon;r,t=p_0+2)=0,$$

and Eqs. (7.97) or (7.98) become (setting $t = p_0 + 1$)

$$\left[\xi^{-1}(\xi^2-\omega^2)^{\frac{1}{2}}\left(\frac{\partial}{\partial\omega}-p_0\omega^{-1}-\frac{\omega(\xi^2-\omega^2)^{-1}}{2}\right)+p_0\omega^{-1}\right]C(\omega,\xi;r,t=p_0)=0, \tag{7.99}$$

which can be solved immediately:

$$C(\omega, \xi; r, t = p_0) = (\xi^2 - \omega^2)^{-\frac{1}{4}}[\xi + (\xi^2 - \omega^2)^{\frac{1}{2}}]^{p_0} h(\xi; r, t = p_0). \tag{7.100}$$

Eliminating $C(\omega, \xi; r, t-1)$ from Eqs. (7.97) and (7.98), setting $t = p_0$, and substituting Eq. (7.100) into the result, we get

$$\left(\frac{d^2}{d\xi^2} + 2(2p_0+1)\xi^{-1}\frac{d}{d\xi} + \xi^{-2}[(2p_0+\tfrac{1}{2})^2 + \rho^2] + 2p_0\xi^{-1} - 1\right) h(\xi; r, t = p_0) = 0. \tag{7.101}$$

Letting

$$h(\xi; r, t = p_0) = \xi^{-(2p_0+1)} g(\xi; r, t = p_0), \tag{7.102}$$

we now reduce Eq. (7.101) to the standard form of Whittaker's equation,

$$\left[\frac{d^2}{d\xi^2} + \left(-1 + \frac{2p_0}{\xi} + \frac{\frac{1}{4}+\rho^2}{\xi^2}\right)\right] g(\xi; r, t = p_0) = 0, \tag{7.103}$$

with the solution

$$g(\xi, r, t = p_0) = cW_{p_0, i\rho}(2\xi). \tag{7.104}$$

Collecting all the factors together, we finally have

$$f(\omega, \xi, \epsilon; r, t = p_0) = C(\xi^2 - \epsilon^2)^{-\frac{1}{4}}[\xi + (\xi^2 - \epsilon^2)^{\frac{1}{2}}]^{p_0}(\xi^2 - \omega^2)^{-\frac{1}{4}} \times [\xi + (\xi^2 - \omega^2)^{\frac{1}{2}}]^{p_0}\xi^{-(2p_0+1)} W_{p_0, i\rho}(2\xi). \tag{7.105}$$

The general $f(\omega, \xi, \epsilon; r, t)$ can be easily obtained through the differential equations (7.97) and (7.98).

So far we have only studied the corner state $|j = p_0, \mu = p_0; k = 0, \nu = 0; \lambda = p_0\rangle$ on the right boundary $p = \lambda$. For an arbitrary state

$$|j = p_0 + s/2, \mu = j; k = s/2, \nu = k; \lambda = p_0 + s\rangle,$$

$s = 0, 1, 2, \cdots$, on the right boundary, we have the following four relations [Eqs. (I.2.41)]:

$$P_+ |j = p_0 + s/2, \mu = j; k = s/2, \nu = k; \lambda = p_0 + s\rangle = (2p_0 + s + 1)^{\frac{1}{2}} a_1(p_0 + s/2, s/2, p_0 + s) \times |j = p_0 + (s+1)/2, \mu = j; k = (s+1)/2, \nu = (s-1)/2; \lambda = p_0 + s + 1\rangle, \tag{7.106a}$$

$$Q_+ |j = p_0 + s/2, \mu = j; k = s/2, \nu = k; \lambda = p_0 + s\rangle = -(s+1)^{\frac{1}{2}} a_1(p_0 + s/2, s/2, p_0 + s) \times |j = p_0 + (s+1)/2, \mu = p_0 + (s-1)/2, k = (s+1)/2, \nu = k; \lambda = p_0 + s + 1\rangle, \tag{7.106b}$$

$$S_+ |j = p_0 + s/2, \mu = j; k = s/2, \nu = k; \lambda = p_0 + s\rangle = -(2p_0 + s + 1)^{\frac{1}{2}}(s+1)^{\frac{1}{2}} a_1(p_0 + s/2, s/2, p_0 + s) \times |j = p_0 + (s+1)/2, \mu = j; k = (s+1)/2, \nu = k; \lambda = p_0 + s + 1\rangle, \tag{7.106c}$$

$$T_+ |j = p_0 + s/2, \mu = j, k = s/2, \nu = k; \lambda = p_0 + s\rangle = a_1(p_0 + s/2, s/2, p_0 + s) \times |j = p_0 + (s+1)/2, \mu = p_0 + (s-1)/2; k = (s+1)/2, \nu = (s-1)/2; \lambda = p_0 + s + 1\rangle, \tag{7.106d}$$

where [Eq. (II.1.7a)]

$$a_1^2\left(p_0 + \frac{s}{2}, \frac{s}{2}, p_0 + s\right) = -\frac{(p_0 + s + \frac{3}{2})^2 + \rho^2}{(s+2)(2p_0 + s + 2)}, \tag{7.107}$$

a negative quantity [Eq. (I.6.4)]. The general matrix element

$$\langle j = p_0 + s/2, \mu = j; k = s/2, \nu = k, \lambda = p_0 + s \mid \omega, \xi, \epsilon; r, t, m = p_0 + s\rangle$$

can now be obtained by the repeated use of Eqs. (7.106c), (7.39)–(7.46), and (7.105). The results, however, are omitted here.

Similarly, matrix elements

$$\langle j, \mu; k, \nu; \lambda \mid \omega, \xi, \epsilon; r, t, m = \mu + \nu\rangle$$

in the interior of $p - \lambda \geq 0$ and $p + \lambda \geq 0$, $p = j + k$, can be obtained in a straightforward manner by using repeatedly Eqs. (I.2.41), (II.1.7), (II.1.8), and (7.85). The mathematics involved is elementary but tedious; we need only do differentiations and solve algebraic equations. The results are again omitted.

2. *The Complementary Series, Class XIV*

The three Casimir operators are

$$c_2 = p_0^2 + 2(\sigma - 1)(\sigma + 2), \tag{7.108a}$$

$$c_3 = \pm p_0\sigma(\sigma + 1), \tag{7.108b}$$

$$c_4 = \tfrac{1}{4}p_0^4 - p_0^2(\sigma^2 + \sigma + 1), \tag{7.108c}$$

where $p_0 = 0, 1, 2, \cdots$, $-1 < \sigma < 0$. Comparing with Eqs. (4.29)–(4.31), we have

$$A = \pm p_0, \tag{7.109a}$$

$$B = \sigma, \tag{7.109b}$$

$$C = B + 1 = \sigma + 1. \tag{7.109c}$$

The allowed states of $SU(2) \times SU(2) \times U(1)$ obey the restriction $p - \lambda \geq 0$ and $p + \lambda \geq 0$. Equations (7.108) are similar to Eqs. (7.91) and can be regarded as analytic continuations of the latter. In fact, from Sec. 5, we know that, for the *even* continuous

series C_q^0 of $SU(1,1)$, $q = -(r-\frac{1}{2})^2 + \frac{1}{4} > 0$, $m = \pm$ integer; hence,

$$(r-\tfrac{1}{2})^2 < \tfrac{1}{4}.$$

Now, we have two possibilities: Either $(r-\frac{1}{2})^2$ is negative and

$$r = \tfrac{1}{2} \pm i\rho \tag{7.110}$$

with ρ real, or $(r-\frac{1}{2})^2$ is positive and

$$0 < r < 1. \tag{7.111}$$

The first possibility, Eq. (7.110), has already been included in our discussion of the principal series. [See the discussion after Eq. (7.93).] The second possibility, Eq. (7.111), is precisely what is needed for the complementary series of $SU(2,2)$ here, where $p_0 = 0, 1, 2, \cdots$, are integers.

If we define the matrix element

$$\langle j = p_0, \mu = p_0; k = 0, \nu = 0; \lambda = p_0 \mid \omega, \xi, \epsilon; r = \sigma + 1, t = p_0, m = p_0\rangle \equiv f(\omega, \xi, \epsilon; r, t = p_0), \tag{7.112}$$

we can find it immediately from Eq. (7.105) by substituting $i\rho \to \sigma + \frac{1}{2}$:

$$f(\omega, \xi, \epsilon; r, t = p_0) = C(\xi^2 - \epsilon^2)^{-\frac{1}{4}}[\xi + (\xi^2 - \epsilon^2)^{\frac{1}{2}}]^{p_0}(\xi^2 - \omega^2)^{-\frac{1}{4}} \times [\xi + (\xi^2 - \omega^2)^{\frac{1}{2}}]^{p_0}\xi^{-(2p_0+1)}W_{p_0,\sigma+\frac{1}{2}}(2\xi). \tag{7.113}$$

3. *The Most Degenerate Discrete Series, p = Half-Integer, Classes V and VI*

The three Casimir operators are (for both Classes V and VI)

$$c_2 = p_0^2 - \tfrac{9}{2}, \tag{7.114a}$$

$$c_3 = \pm\tfrac{1}{4}p_0, \tag{7.114b}$$

$$c_4 = \tfrac{1}{4}p_0^4 - \tfrac{3}{4}p_0^2, \tag{7.114c}$$

where $p_0 = \frac{1}{2}, \frac{3}{2}, \frac{5}{2}, \cdots$. Equations (7.114) are the $\rho = 0$ limits of Eqs. (7.90). But the $\rho = 0$ limit of the principal series (the half-integer p_0 part) splits up into *two* inequivalent series of unitary irreducible representations.

(*a*) *Class V:* The allowed states of the maximal compact subgroup $SU(2) \times SU(2) \times U(1)$ obey the restriction $p - \lambda \geq 0$ and $p + \lambda \geq p_0 + \frac{1}{2}$ [see Eqs. (II.3.16) and (II.3.18)]. We again specialize to $c_3 = \frac{1}{4}p_0$, and consider the right corner state

$$|j = p_0, \mu = p_0; k = 0, \nu = 0; \lambda = p_0\rangle.$$

Let

$$\langle j = p_0, \mu = p_0; k = 0, \nu = 0; \lambda = p_0 \mid \omega, \xi, \epsilon; r = \tfrac{1}{2}, t, m = p_0\rangle \equiv f(\omega, \xi, \epsilon; t). \tag{7.115}$$

The special case with $t = p_0$ can be written down immediately from Eq. (7.105), with $\rho = 0$:

$$f(\omega, \xi, \epsilon; t = p_0) = c(\xi^2 - \epsilon^2)^{-\frac{1}{4}}[\xi + (\xi^2 - \epsilon^2)^{\frac{1}{2}}]^{p_0}(\xi^2 - \omega^2)^{-\frac{1}{4}} \times [\xi + (\xi^2 - \omega^2)^{\frac{1}{2}}]^{p_0}\xi^{-(2p_0+1)}W_{p_0,0}(2\xi), \tag{7.116}$$

where the Whittaker function of the second kind $W_{p_0,0}(2\xi)$ can be expressed in terms of the Laguerre polynomials:

$$W_{p_0,0}(2\xi) = (-1)^{p_0-\frac{1}{2}}(p_0 - \tfrac{1}{2})!\, e^{-\xi}(2\xi)^{\frac{1}{2}}L_{p_0-\frac{1}{2}}(2\xi). \tag{7.117}$$

The general $f(\omega, \xi, \epsilon; t \neq p_0)$ can again be obtained easily through Eqs. (7.97) and (7.98) by setting $r = \frac{1}{2}$. However, when $t = \pm\frac{1}{2}$, $c(\omega, \xi; r = \frac{1}{2}, t = \pm\frac{1}{2})$ are decoupled from each other. This interesting result is a direct reflection of the recurrence relation for Whittaker functions

$$(2\kappa - z)W_{\kappa,\mu}(z) + W_{\kappa+1,\mu}(z) = (\mu - \kappa + \tfrac{1}{2})(\mu + \kappa - \tfrac{1}{2})W_{\kappa-1,\mu}(z) \tag{7.118}$$

and agrees with the discussion in Sec. 5, where we remarked that for the subgroup $SU(1,1)$ its D^+ series is inequivalent to its D^- series. For $r = \frac{1}{2}$ the UIR in the D^+ series requires $t = \frac{1}{2}, \frac{3}{2}, \frac{5}{2}, \cdots$, while the UIR in the D^- series requires $t = -\frac{1}{2}, -\frac{3}{2}, -\frac{5}{2}, \cdots$; the two parts are completely decoupled. Therefore, knowing the solution $f(\omega, \xi, \epsilon; t = p_0)$ from Eqs. (7.116) and (7.117) does not give us any information about $f(\omega, \xi, \epsilon; t \leq -\frac{1}{2})$. To find out the properties of the latter, we consider the function $f(\omega, \xi, \epsilon; t = -p_0)$, which can be obtained in the same way as Eq. (7.105):

$$f(\omega, \xi, \epsilon; t = -p_0) = c'(\xi^2 - \epsilon^2)^{-\frac{1}{4}}[\xi - (\xi^2 - \epsilon^2)^{\frac{1}{2}}]^{p_0}(\xi^2 - \omega^2)^{-\frac{1}{4}} \times [\xi - (\xi^2 - \omega^2)^{\frac{1}{2}}]^{p_0}\xi^{-(2p_0+1)}W_{-p_0,0}(2\xi). \tag{7.119}$$

To simplify the discussion and concentrate on the salient features of the problem, let us study the special case $p_0 = \frac{1}{2}$; then $t = \pm\frac{1}{2}$ only. Equation (7.98) is now reduced to (remember $r = \frac{1}{2}$)

$$\left(\frac{\omega}{\xi}\frac{\partial}{\partial\omega} + \frac{\partial}{\partial\xi} + \frac{3}{2}\frac{1}{\xi} + 1\right)c\left(\omega, \xi; t = \frac{1}{2}\right) = 0, \tag{7.120a}$$

$$\left(\frac{\omega}{\xi}\frac{\partial}{\partial\omega} + \frac{\partial}{\partial\xi} + \frac{3}{2}\frac{1}{\xi} - 1\right)c\left(\omega, \xi; t = -\frac{1}{2}\right) = 0. \tag{7.120b}$$

The solution for Eq. (7.20a) is

$$c(\omega, \xi; t = \tfrac{1}{2}) = c(\xi^2 - \omega^2)^{-\frac{1}{4}}[\xi + (\xi^2 - \omega^2)^{\frac{1}{2}}]^{\frac{1}{2}}\xi^{-\frac{3}{2}}e^{-\xi}, \tag{7.121a}$$

which agrees with Eqs. (7.116) and (7.117) as expected. However, the solution for Eq. (7.120b) is

$$c(\omega, \xi; t = -\tfrac{1}{2}) = c'(\xi^2 - \omega^2)^{-\frac{1}{4}}[\xi - (\xi^2 - \omega^2)^{\frac{1}{2}}]^{\frac{1}{2}}\xi^{-\frac{3}{2}}e^{\xi}, \quad (7.121b)$$

which diverges as $\xi \to \infty$. Equation (7.121b) is compatible with Eq. (7.119) if instead of $W_{-p_0,0}(2\xi)$ we had $M_{-p_0,0}(2\xi)$, which is a Whittaker function of the first kind and diverges at ∞. Therefore, for proper behavior at infinity, we have to set

$$f(\omega, \xi, \epsilon; t = -p_0) = 0, \quad (7.122)$$

and the states $|\omega, \xi, \epsilon; r, t \leq -\tfrac{1}{2}, m = p_0\rangle$ do *not* exist.

(*b*) *Class VI:* The allowed states of the maximal compact subgroup $SU(2) \times SU(2) \times U(1)$ obey the restriction $p + \lambda \geq 0$ and $p - \lambda \geq p_0 + \tfrac{1}{2}$. [See Eqs. (II.3.17) and (II.3.19)]. We again specialize to $c_3 = \tfrac{1}{4}p_0$, and consider the *left* corner state

$$|j = 0, \mu = 0; k = p_0, \nu = p_0; \lambda = -p_0\rangle.$$

Let

$$\langle j = 0, \mu = 0; k = p_0, \nu = p_0; \lambda = -p_0 \,|\, \omega, \xi, \epsilon; r = \tfrac{1}{2}, t, m = p_0\rangle = \tilde{f}(\omega, \xi, \epsilon; t). \quad (7.123)$$

Equation (7.41) now gives us

$$\langle j = 0, \mu = 0; k = p_0, \nu = p_0; \lambda = -p_0| \times \mathfrak{R}_5 - \mathfrak{I}_5 - 2p_0 |\omega, \xi, \epsilon; r = \tfrac{1}{2}, t, m = p_0\rangle = 0. \quad (7.124)$$

Following the same steps as that of Eqs. (7.95)–(7.102) and remembering that Eq. (7.124) differs from Eq. (7.92) only in the sign of the last term, we get (for $t = -p_0$)

$$\tilde{f}(\omega, \xi, \epsilon; t = -p_0) = \tilde{c}(\xi^2 - \epsilon^2)^{-\frac{1}{4}}[\xi - (\xi^2 - \epsilon^2)^{\frac{1}{2}}]^{p_0}(\xi^2 - \omega^2)^{-\frac{1}{4}} \times [\xi - (\xi^2 - \omega^2)^{\frac{1}{2}}]^{p_0}\xi^{-(2p_0+1)}W_{p_0,0}(2\xi). \quad (7.125)$$

Similarly, we find that

$$\tilde{f}(\omega, \xi, \epsilon; t = p_0) = 0, \quad (7.126)$$

and the states $|\omega, \xi, \epsilon; r = \tfrac{1}{2}, t \geq \tfrac{1}{2}, m = p_0\rangle$ do *not* exist.

For an arbitrary state

$$|j = s/2, \mu = j; k = p_0 + s/2, \nu = k; \lambda = -p_0 - s\rangle,$$

$s = 0, 1, 2, \cdots$, on the left boundary, we may use the following relations [Eqs. (I.2.43)]:

$$P_-|j = s/2, \mu = j; k = p_0 + s/2, \nu = k; \lambda = -p_0 - s\rangle = -(2p_0 + s + 1)^{\frac{1}{2}}b_1(s/2, p_0 + s/2, -p_0 - s) \times |j = (s+1)/2, \mu = (s-1)/2; k = p_0 + (s+1)/2, \nu = k; \lambda = -p_0 - s - 1\rangle \quad (7.127)$$

etc., where [Eq. (II.1.8a)]

$$b_1^2\left(\frac{s}{2}, p_0 + \frac{s}{2}, -p_0 - s\right) = -\frac{(p_0 + s + \frac{3}{2})^2}{(s+2)(2p_0 + s + 2)}. \quad (7.128)$$

4. *The Most Degenerate Discrete Series, p = Integer, Classes VII–IX*

The three Casimir operators are (for all three classes VII–IX)

$$c_2 = p_0^2 - 4, \quad (7.129a)$$

$$c_3 = 0, \quad (7.129b)$$

$$c_4 = \tfrac{1}{4}p_0^4 - p_0^2, \quad (7.129c)$$

where $p_0 = 1, 2, 3, \cdots$. Equations (7.129) are the $\sigma = 0$ limits of Eqs. (7.108), but the $\sigma = 0$ limit of the complementary series (the $p_0 = 1, 2, 3, \cdots$ part) splits up into *three* inequivalent series of unitary irreducible representations.

(*a*) *Class VII:* The allowed states of the subgroup $SU(2) \times SU(2) \times U(1)$ obey the restriction $p - \lambda \geq 0$ and $p + \lambda \geq p_0 + 1$ [see Eqs. (II.3.21) and (II.3.23)]. We again consider the right corner state $|j = p_0, \mu = p_0, k = 0, \nu = 0; \lambda = p_0\rangle$. Let

$$\langle j = p_0, \mu = p_0; k = 0, \nu = 0; \lambda = p_0 \,|\, \omega, \xi, \epsilon; r = 1, t, m = p_0\rangle \equiv f(\omega, \xi, \epsilon; t). \quad (7.130)$$

The special case with $t = p_0$ again can be written down immediately from Eq. (7.113) with $\sigma = 0$:

$$f(\omega, \xi, \epsilon; t = p_0) = c(\xi^2 - \epsilon^2)^{-\frac{1}{4}}[\xi + (\xi^2 - \epsilon^2)^{\frac{1}{2}}]^{p_0}(\xi^2 - \omega^2)^{-\frac{1}{4}} \times [\xi + (\xi^2 - \omega^2)^{\frac{1}{2}}]^{p_0}\xi^{-(2p_0+1)}W_{p_0,\frac{1}{2}}(2\xi). \quad (7.131)$$

The general $f(\omega, \xi, \epsilon; t \neq p_0)$ can again be obtained through Eqs. (7.97) and (7.98) by setting $r = 1$. However, we notice that, when $t = 0, \pm 1$, $c(\omega, \xi; r = 1, t = 0, \pm 1)$ are decoupled from each other. An analysis similar to that of Eqs. (7.119)–(7.121) demonstrates that the states $|\omega, \xi, \epsilon; r = 1, t \leq 0, m = p_0\rangle$ do *not* exist. This observation is but a reflection of the fact that here we are using a UIR in

the D^+ series of $SU(1, 1)$ with $r = 1$, and only $t = r$, $r + 1, r + 2, \cdots$ are allowed (see Sec. 5).

(*b*) *Class VIII:* The allowed states of $SU(2) \times SU(2) \times U(1)$ obey the restriction $p + \lambda \geq 0$, $p - \lambda \geq p_0 + 1$ [see Eqs. (II.3.22) and (II.3.24)]. We now consider the *left* corner state $|j = 0, \mu = 0; k = p_0, \nu = p_0; \lambda = -p_0\rangle$. Following the same procedure as in Eqs. (7.124) and (7.125), we now get

$$\tilde{f}(\omega, \xi, \epsilon; t = -p_0) = \tilde{c}(\xi^2 - \epsilon^2)^{-\frac{1}{4}}[\xi - (\xi^2 - \epsilon^2)^{\frac{1}{4}}]^{p_0}(\xi^2 - \omega^2)^{-\frac{1}{4}} \times [\xi - (\xi^2 - \omega^2)^{\frac{1}{2}}]^{p_0}\xi^{-(2p_0+1)}W_{p_0,\frac{1}{2}}(2\xi), \quad (7.132)$$

and the states $|\omega, \xi, \epsilon; r = 1, t \geq 0, m = p_0\rangle$ do *not* exist. Here, we are using a UIR in the D^- series of $SU(1, 1)$ with $r = 1$, and only $t = -r, -(r + 1), \cdots$ are allowed.

(*c*) *Class IX:* The allowed states of $SU(2) \times SU(2) \times U(1)$ obey the restriction $p - \lambda \geq p_0$ and $p + \lambda \geq p_0$ [see Eq. (II.3.25)]. We now consider the "ground" state

$$|j = p_0/2, \mu = p_0/2; k = p_0/2, \nu = p_0/2; \lambda = 0\rangle.$$

Let

$$\langle j = p_0/2, \mu = p_0/2; k = p_0/2, \nu = p_0/2; \lambda = 0 \,|\, \omega, \xi, \epsilon, r = 1, t = 0, m = p_0\rangle = f(\omega, \xi, \epsilon). \quad (7.133)$$

Equation (7.41) now gives

$$\langle j = p_0/2, \mu = p_0/2; k = p_0/2, \nu = p_0/2; \lambda = 0| \times \mathfrak{R}_5 - \mathfrak{I}_5 |\omega, \xi, \epsilon; r = 1, t = 0, m = p_0\rangle = 0, \quad (7.134)$$

and Eq. (7.97) leads to

$$\left(\frac{\partial^2}{\partial\omega^2} + \frac{\partial^2}{\partial\xi^2} + 2\frac{\omega}{\xi}\frac{\partial}{\partial\omega}\frac{\partial}{\partial\xi} + \frac{1}{\omega}\frac{\partial}{\partial\omega} + \frac{3}{\xi}\frac{\partial}{\partial\xi} - p_0^2\frac{1}{\omega^2} - 1 - \frac{(\xi^2 - \omega^2)^{-1}}{4}\right)c(\omega, \xi) = 0, \quad (7.135)$$

which can be solved easily:

$$c(\omega, \xi) = c(\xi^2 - \omega^2)^{-\frac{1}{4}}\omega^{p_0}\xi^{-(p_0+\frac{1}{2})}K_{p_0+\frac{1}{2}}(\xi). \quad (7.136)$$

Therefore,

$$f(\omega, \xi, \epsilon) = c\epsilon^{p_0}(\xi^2 - \epsilon^2)^{-\frac{1}{4}}\omega^{p_0}(\xi^2 - \omega^2)^{-\frac{1}{4}}\xi^{-(p_0+\frac{1}{2})}K_{p_0+\frac{1}{2}}(\xi), \quad (7.137)$$

where $K_{p_0+\frac{1}{2}}(\xi)$ is the modified Bessel's function of the second kind. It can be shown that $f(\omega, \xi, \epsilon; t \neq 0) = 0$ and the states $|\omega, \xi, \epsilon; r = 1, t \neq 0, m = p_0\rangle$ do *not* exist. Here, the UIR of $SU(1, 1)$ involved is the *trivial* 1-dimensional unitary irreducible representation.

5. The Most Degenerate Discrete Representation, $p_0 = 0$, Class X

This is an isolated representation; the three Casimir operators are

$$c_2 = -4, \quad (7.138a)$$

$$c_3 = 0, \quad (7.138b)$$

$$c_4 = 0. \quad (7.138c)$$

This representation is the $\sigma = 0$ limit of the complementary series with $p_0 = 0$. The allowed states of $SU(2) \times SU(2) \times U(1)$ obey the restriction $p - \lambda \geq 0$ and $p + \lambda \geq 0$ [see Eq. (II.3.20)]. The "ground state" is $|j = 0, \mu = 0; k = 0, \nu = 0; \lambda = 0\rangle$. The transformation function

$$f(\omega, \xi, \epsilon) = \langle j = 0, \mu = 0; k = 0, \nu = 0; \lambda = 0 \,|\, \omega, \xi, \epsilon; r = 1, t = 0, m = 0\rangle \quad (7.139)$$

is a special case of Eq. (7.133) with $p_0 = 0$, and the solution is

$$f(\omega, \xi, \epsilon) = c(\xi^2 - \epsilon^2)^{-\frac{1}{4}}(\xi^2 - \omega^2)^{-\frac{1}{4}}\xi^{-\frac{1}{2}}K_{\frac{1}{2}}(\xi) = c'(\xi^2 - \epsilon^2)^{-\frac{1}{4}}(\xi^2 - \omega^2)^{-\frac{1}{4}}\xi^{-1}e^{-\xi}. \quad (7.140)$$

8. CONCLUSION AND SUMMARY

We have come to the end of our study of the reduction of $SU(2, 2)$ with respect to the iso-Poincaré subgroup $E(3, 1)$. The major calculations completed here are the construction of the representations of $\mathfrak{R}_\mu$, the conformal operators, in Eqs. (4.11)–(4.14) and (6.18)–(6.21), and the eigenvalues of the three Casimir operators, Eqs. (4.19)–(4.24). These calculations were followed in Sec. 7 by the computation of the matrix elements between a state classified by the maximal compact subgroup $SU(2) \times SU(2) \times U(1)$, $|j, \mu; k, \nu; \lambda\rangle$ and a state classified by $E(3, 1)$, $|\eta, \xi, \epsilon; s, t, m\rangle$, $|\omega, \xi, \epsilon; r, t, m\rangle$, or $|\xi, \epsilon; t, m\rangle$. We found that the fourteen classes of unitary irreducible representations of $SU(2, 2)$ are divided into four distinct groups: (1) Two classes are reducible with respect to the "lightlike" UIR's of $E(3, 1)$ only; (2) two classes are reducible with respect to the "timelike" UIR's of $E(3, 1)$ only; (3) eight classes are reducible with respect to the "spacelike" UIR's of $E(3, 1)$ only; (4) two classes contain both "timelike" and "spacelike" UIR's of $E(3.1)$. Once the matrix elements involving the "ground" states or states of highest weights have been calculated, the general matrix elements can be obtained directly by simple algebra and repeated differentiations.

APPENDIX A: RELATION BETWEEN L_{ab} AND A_α^β

The canonical generators A_α^β of $SU(4)$ satisfy the commutation relations (2, 1). The generators of the maximal compact subgroup $SU(2) \times SU(2) \times U(1)$ of $SU(2,2)$ are written as

$$J_+ = A_1^2, \quad J_- = A_2^1, \quad J_3 = \tfrac{1}{2}(A_1^1 - A_2^2), \tag{A1}$$

$$K_+ = A_3^4, \quad K_- = A_4^3, \quad K_3 = \tfrac{1}{2}(A_3^3 - A_4^4), \tag{A2}$$

$$R_0 = \tfrac{1}{2}(A_1^1 + A_2^2 - A_3^3 - A_4^4). \tag{A3}$$

The remaining eight generators are

$$P_+ = iA_1^3, \quad P_- = iA_3^1, \tag{A4}$$

$$Q_+ = iA_2^4, \quad Q_- = iA_4^2, \tag{A5}$$

$$S_+ = iA_1^4, \quad S_- = iA_4^1, \tag{A6}$$

$$T_+ = iA_2^3, \quad T_- = iA_3^2. \tag{A7}$$

The relation between L_{ab}, the canonical generators of $O(4,2)$, and the generators of $SU(2,2)$ can be written down as follows:

$$J_\pm = \tfrac{1}{2}[(L_{23} + L_{15}) \pm i(L_{31} + L_{25})], \tag{A8}$$

$$J_3 = \tfrac{1}{2}(L_{12} + L_{35}), \tag{A9}$$

$$K_\pm = \tfrac{1}{2}[(L_{23} - L_{15}) \pm (L_{31} - L_{25})], \tag{A10}$$

$$K_3 = \tfrac{1}{2}(L_{12} - L_{35}), \tag{A11}$$

$$R_0 = L_{06}, \tag{A12}$$

$$P_\pm = \tfrac{1}{2}[(L_{56} + L_{03}) \pm i(L_{05} + L_{36})], \tag{A13}$$

$$Q_\pm = \tfrac{1}{2}[(L_{56} - L_{03}) \pm i(L_{05} - L_{36})], \tag{A14}$$

$$S_\pm = \tfrac{1}{2}(L_{0\pm} \pm iL_{\pm 6}), \tag{A15}$$

$$T_\pm = \tfrac{1}{2}(L_{0\mp} \pm iL_{\mp 6}), \tag{A16}$$

where

$$L_{\pm 6} = L_{16} \pm iL_{26} \quad \text{and} \quad L_{0\pm} = L_{01} \pm iL_{02}.$$

Next, we study the relation between the 15 Dirac γ matrices and L_{ab}. We have immediately

$$\gamma_\mu \sim 2L_{\mu 6}, \tag{A17}$$

$$\sigma_{\mu\nu} \sim 2L_{\mu\nu}, \tag{A18}$$

$$i\gamma_5\gamma_\mu \sim 2L_{\mu 5}, \tag{A19}$$

$$\gamma_5 \sim 2L_{56}, \tag{A20}$$

where $\mu, \nu = 0, 1, 2, 3$ and $\sim$ means "transforms like." Our γ matrices are so chosen that

$$\{\gamma_\mu, \gamma_\nu\} = -2g_{\mu\nu}, \tag{A21}$$

$$\sigma_{\mu\nu} = \tfrac{1}{2}i[\gamma_\mu, \gamma_\nu], \tag{A22}$$

$$\gamma_5 = \gamma_1\gamma_2\gamma_3\gamma_0, \tag{A23}$$

with $g_{11} = g_{22} = g_{33} = -g_{00} = 1$. The factor in Eqs. (A17)–(A20) is included so that in the 4-dimensional representation the "transform like" sign $\sim$ becomes an equality sign. For completeness we write down the explicit representation for the Dirac matrices,

$$\boldsymbol{\gamma} = \begin{pmatrix} 0 & \boldsymbol{\sigma} \\ -\boldsymbol{\sigma} & 0 \end{pmatrix}, \quad \gamma_0 = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix},$$

$$\sigma_{ij} = \epsilon_{ijk}\begin{pmatrix} \sigma_k & 0 \\ 0 & \sigma_k \end{pmatrix}, \quad \sigma_{0k} = i\begin{pmatrix} 0 & \sigma_k \\ \sigma_k & 0 \end{pmatrix},$$

$$i\gamma_5\boldsymbol{\gamma} = \begin{pmatrix} \boldsymbol{\sigma} & 0 \\ 0 & -\boldsymbol{\sigma} \end{pmatrix}, \quad i\gamma_5\gamma_0 = \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix},$$

$$\gamma_5 = i\begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, \tag{A24}$$

where $i, j = 1, 2, 3$ and $\boldsymbol{\sigma}$ are the Pauli matrices. We notice that γ_0, σ_{ij}, and $i\gamma_5\boldsymbol{\gamma}$ are Hermitian, while $\boldsymbol{\gamma}$, σ_{0k}, $i\gamma_5\gamma_0$, and γ_5 are anti-Hermitian. This choice is made because the Hermitian matrices are linear combinations of $\mathbf{J}$, $\mathbf{K}$, and R_0 only, involving generators of the maximal compact subgroup, while the anti-Hermitian matrices are linear combinations of $P_\pm$, $Q_\pm$, $S_\pm$, and $T_\pm$ only, involving the noncompact generators of $SU(2,2)$.

APPENDIX B: DERIVATION FOR $\mathfrak{L}_{\pm 6}$ AND $\mathfrak{L}_{56}$

We start with the representation for $\mathfrak{L}_{56}$. The most general expression is

$$\mathfrak{L}_{56}\,|\eta, \xi, \epsilon; s, t, m\rangle = -i\sum_{n=-1}^{1}\left(g_n\frac{\partial}{\partial\xi} + h_n\frac{\partial}{\partial\epsilon} + f_n(m)\right) \times |\eta, \xi, \epsilon; s, t+n, m\rangle, \tag{B1}$$

where $n = -1, 0, 1$ and g_n, h_n, and f_n are functions of η, ξ, and ϵ. From

$$[\mathfrak{J}_\pm, \mathfrak{L}_{56}] = 0, \tag{B2}$$

we get

$$h_n + \epsilon f_n(m) = \epsilon f_n(m \pm 1), \tag{B3}$$

and f_n is independent of m, with

$$h_n = 0. \tag{B4}$$

Next, we consider the commutation relation

$$[\mathfrak{J}_5, \mathfrak{L}_{56}] = -i\mathfrak{J}_6, \tag{B5}$$

which leads to

$$g_\pm = 0 \tag{B6}$$

and

$$g_0 = \xi^{-1}(\xi^2 + \eta^2)^{\frac{1}{2}}(\xi^2 - \epsilon^2)^{\frac{1}{2}}. \tag{B7}$$

The commutation relation $[\mathfrak{J}_6, \mathfrak{L}_{56}] = -i\mathfrak{J}_5$ is now automatically satisfied.

We can now rewrite Eq. (B1) as follows:

$$\begin{aligned}\mathfrak{L}_{56}|\eta,\xi,\epsilon;s,t,m\rangle &= -i\left(\xi^{-1}(\xi^2+\eta^2)^{\frac{1}{2}}(\xi^2-\epsilon^2)^{\frac{1}{2}}\frac{\partial}{\partial\xi}+f_0\right)\\ &\times|\eta,\xi,\epsilon;s,t,m\rangle\\ &- if_+|\eta,\xi,\epsilon;s,t+1,m\rangle\\ &- if_-|\eta,\xi,\epsilon;s,t-1,m\rangle,\end{aligned}\tag{B8}$$

where the three functions f_n, $n=-1, 0, 1$, are yet to be determined. Next, we consider

$$[\mathfrak{L}_{56},\mathfrak{L}_+]=\mathfrak{L}_{+6}\tag{B9}$$

and obtain

$$\begin{aligned}\mathfrak{L}_{+6}|\eta,\xi,\epsilon;s,t,m\rangle &= -i\Big[(\xi^2+\eta^2)^{\frac{1}{2}}\Big(\frac{\epsilon}{\xi}\frac{\partial}{\partial\xi}+\frac{\partial}{\partial\epsilon}+(m+1)\epsilon^{-1}\\ &+\frac{\epsilon(\xi^2-\epsilon^2)^{-1}}{2}-t\xi^{-1}\epsilon^{-1}(\xi^2-\epsilon^2)^{\frac{1}{2}}\Big)\\ &-(\xi^2-\epsilon^2)^{\frac{1}{2}}\frac{\partial}{\partial\epsilon}f_0\Big]|\eta,\xi,\epsilon;s,t,m+1\rangle\\ &+i\Big((\xi^2-\epsilon^2)^{\frac{1}{2}}\frac{\partial}{\partial\epsilon}+\xi\epsilon^{-1}\Big)\\ &\times f_+|\eta,\xi,\epsilon;s,t+1,m+1\rangle\\ &+i\Big((\xi^2-\epsilon^2)^{\frac{1}{2}}\frac{\partial}{\partial\epsilon}-\xi\epsilon^{-1}\Big)\\ &\times f_-|\eta,\xi,\epsilon;s,t-1,m+1\rangle.\end{aligned}\tag{B10}$$

From

$$[\mathfrak{L}_+,\mathfrak{L}_{+6}]=0,\tag{B11}$$

we get

$$(\xi^2-\epsilon^2)\frac{\partial^2}{\partial\epsilon^2}f_0-\xi^2\epsilon^{-1}\frac{\partial}{\partial\epsilon}f_0-\epsilon^2(\xi^2+\eta^2)^{\frac{1}{2}}(\xi^2-\epsilon^2)^{-\frac{3}{2}}=0\tag{B12}$$

and

$$f_\pm=\mp\epsilon d_\pm,\tag{B13}$$

where the functions $d_\pm$ are independent of ϵ. Now, we restrict ourselves to unitary irreducible representations, and

$$(f(\eta',\xi',\epsilon';s,t,m),\mathfrak{L}_{56}f(\eta,\xi,\epsilon;s,t,m))=(\mathfrak{L}_{56}f(\eta',\xi',\epsilon';s,t,m),f(\eta,\xi,\epsilon;s,t,m)),\tag{B14}$$

which upon using (B8) leads to

$$f_0=\tfrac{1}{2}(\xi^2+\eta^2)^{-\frac{1}{2}}(\xi^2-\epsilon^2)^{\frac{1}{2}}+\tfrac{1}{2}(\xi^2+\eta^2)^{\frac{1}{2}}(\xi^2-\epsilon^2)^{-\frac{1}{2}}\tag{B15}$$

and

$$d_+(t)=d_-(t+1)\equiv d(t+1).\tag{B16}$$

We rewrite Eqs. (B8) and (B10) as follows:

$$\begin{aligned}\mathfrak{L}_{56}|\eta,\xi,\epsilon;s,t,m\rangle &= -i\Big((\xi^2+\eta^2)^{\frac{1}{2}}(\xi^2-\epsilon^2)^{\frac{1}{2}}\xi^{-1}\frac{\partial}{\partial\xi}\\ &+\frac{(\xi^2+\eta^2)^{-\frac{1}{2}}(\xi^2-\epsilon^2)^{\frac{1}{2}}}{2}+\frac{(\xi^2+\eta^2)^{\frac{1}{2}}(\xi^2-\epsilon^2)^{-\frac{1}{2}}}{2}\Big)\\ &\times|\eta,\xi,\epsilon;s,t,m\rangle\\ &+i\epsilon d(t+1)|\eta,\xi,\epsilon;s,t+1,m\rangle\\ &-i\epsilon d(t)|\eta,\xi,\epsilon;s,t-1,m\rangle,\end{aligned}\tag{B17}$$

$$\begin{aligned}\mathfrak{L}_{\pm6}|\eta,\xi,\epsilon;s,t,m\rangle &= -i(\xi^2+\eta^2)^{\frac{1}{2}}\Big(\epsilon\xi^{-1}\frac{\partial}{\partial\xi}+\frac{\partial}{\partial\epsilon}\pm(m\pm1)\epsilon^{-1}\\ &+\frac{\epsilon(\xi^2+\eta^2)^{-1}}{2}\mp t\xi^{-1}\epsilon^{-1}(\xi^2-\epsilon^2)^{\frac{1}{2}}\Big)\\ &\times|\eta,\xi,\epsilon;s,t,m\pm1\rangle\\ &\mp i[\xi\pm(\xi^2-\epsilon^2)^{\frac{1}{2}}]d(t+1)\\ &\times|\eta,\xi,\epsilon;s,t+1,m\pm1\rangle\\ &\mp i[\xi\mp(\xi^2-\epsilon^2)^{\frac{1}{2}}]d(t)\\ &\times|\eta,\xi,\epsilon;s,t-1,m\pm1\rangle,\end{aligned}\tag{B18}$$

where we have also used

$$[\mathfrak{L}_-,\mathfrak{L}_{56}]=\mathfrak{L}_{-6}.\tag{B19}$$

In order to determine the function $d(t)$, we study the covariant spin operator W_μ, with

$$W_\mu W^\mu=\mathbf{W}^2-W_6^2.\tag{B20}$$

The representations for W_μ can be easily obtained from Eqs. (B17) and (B18):

$$\begin{aligned}W_\pm|\eta,\xi,\epsilon;s,t,m\rangle &= -\epsilon\xi^{-1}(\xi^2+\eta^2)^{\frac{1}{2}}t|\eta,\xi,\epsilon;s,t,m\pm1\rangle\\ &\mp\xi[\xi\pm(\xi^2-\epsilon^2)^{\frac{1}{2}}]d(t+1)\\ &\times|\eta,\xi,\epsilon;s,t+1,m\pm1\rangle\\ &\pm\xi[\xi\mp(\xi^2-\epsilon^2)^{\frac{1}{2}}]d(t)\\ &\times|\eta,\xi,\epsilon;s,t-1,m\pm1\rangle,\end{aligned}\tag{B21}$$

$$\begin{aligned}W_5|\eta,\xi,\epsilon;s,t,m\rangle &= -\xi^{-1}(\xi^2+\eta^2)^{\frac{1}{2}}(\xi^2-\epsilon^2)^{\frac{1}{2}}t|\eta,\xi,\epsilon;s,t,m\rangle\\ &+\epsilon\xi d(t+1)|\eta,\xi,\epsilon;s,t+1,m\rangle\\ &+\epsilon\xi d(t)|\eta,\xi,\epsilon;s,t-1,m\rangle,\end{aligned}\tag{B22}$$

which leads to

$$\begin{aligned}\mathbf{W}^2|\eta,\xi,\epsilon;s,t,m\rangle &= [(\xi^2+\eta^2)t^2+2\xi^4d^2(t+1)+2\xi^4d^2(t)]\\ &\times|\eta,\xi,\epsilon;s,t,m\rangle.\end{aligned}\tag{B23}$$

Using Eqs. (3.2), (2.23), and (B20), we have

$$2\xi^4[d^2(t+1)+d^2(t)] = \eta^2[s(s+1)-t^2]. \quad \text{(B24)}$$

Finally, we consider

$$[\mathfrak{L}_{+6}, \mathfrak{L}_{-6}] = -2\mathfrak{L}_5, \quad \text{(B25)}$$

which gives us

$$d^2(t) - d^2(t+1) = \tfrac{1}{2}\eta^2\xi^{-4}t. \quad \text{(B26)}$$

Solving Eqs. (B24) and (B26), we get

$$d(t) = \tfrac{1}{2}\eta\xi^{-2}[(s+t)(s-t+1)]^{\frac{1}{2}} \quad \text{(B27)}$$

and $t = -s, -s+1, \cdots, s$, so that $d(t)$ is real. The final results for $\mathfrak{L}_{\pm 6}$, $\mathfrak{L}_{56}$, and W_μ have been stated in Sec. 3.

APPENDIX C: OUTLINE OF DERIVATION FOR $\mathcal{R}_\mu$

The procedure for the derivation of Eqs. (4.11)–(4.14) is similar to that given in Appendix B, only more complicated. Here, we shall only give an outline of our method. We start with $\mathcal{R}_5$ and write down the most general expression, Eq. (4.10), which involves 90 unknown functions.

From

$$[\mathfrak{I}_5, \mathcal{R}_5] = -2iL_{03}, \quad \text{(C1)}$$

we get

$$\xi D_{00} - \epsilon E_{00} = 2\eta(\xi^2-\epsilon^2)^{\frac{1}{2}}, \quad \text{(C2)}$$

$$2\xi B_{00} - \epsilon F_{00} = 2\xi(\xi^2-\epsilon^2)^{\frac{1}{2}}, \quad \text{(C3)}$$

$$\xi F_{00} - 2\epsilon C_{00} = 2\epsilon(\xi^2-\epsilon^2)^{\frac{1}{2}}, \quad \text{(C4)}$$

$$-\epsilon^2 B_{00} - \xi^2 C_{00} + \epsilon\xi F_{00} + \xi(\xi^2-\epsilon^2)b_{00} - \epsilon(\xi^2-\epsilon^2)c_{00} = 6(\xi^2-\epsilon^2)^{\frac{3}{2}}, \quad \text{(C5)}$$

$$\xi b_{kl} - \epsilon c_{kl} = 0, \quad k, l \text{ not all } 0. \quad \text{(C6)}$$

Next, we consider

$$[\mathfrak{I}_6, \mathcal{R}_5] = 2i\mathfrak{L}_{56}, \quad \text{(C7)}$$

from which we obtain [using Eqs. (C2)–(C6)]

$$A_{kl} = \tfrac{1}{2}\epsilon\xi\eta^{-2}F_{kl} - \delta_{kl}(\xi^2-\epsilon^2)^{\frac{1}{2}}, \quad \text{(C8)}$$

$$B_{kl} = \tfrac{1}{2}\epsilon\xi^{-1}F_{kl} + \delta_{kl}(\xi^2-\epsilon^2)^{\frac{1}{2}}, \quad \text{(C9)}$$

$$C_{kl} = \tfrac{1}{2}\xi\epsilon^{-1}F_{kl} - \delta_{kl}(\xi^2-\epsilon^2)^{\frac{1}{2}}, \quad \text{(C10)}$$

$$D_{kl} = -\epsilon\eta^{-1}F_{kl} + 2\delta_{kl}\eta\xi^{-1}(\xi^2-\epsilon^2)^{\frac{1}{2}}, \quad \text{(C11)}$$

$$E_{kl} = -\xi\eta^{-1}F_{kl}, \quad \text{(C12)}$$

where $k, l = 0, \pm$ and δ_{kl} is the Kronecker delta. We have

$$\eta a_{00} + \xi b_{00} = 2(\xi^2-\epsilon^2)^{\frac{1}{2}} + (\xi^2+\eta^2)(\xi^2-\epsilon^2)^{-\frac{1}{2}} - \tfrac{1}{2}\epsilon\xi^{-1}\eta^{-2}(\xi^2+\eta^2)F_{00}, \quad \text{(C13)}$$

$$\eta a_{0\pm} + \xi b_{0\pm} = \mp\epsilon\eta\xi^{-2}(\xi^2+\eta^2)^{\frac{1}{2}} \times [(s\mp t)(s\pm t+1)]^{\frac{1}{2}}. \quad \text{(C14)}$$

The commutation relations

$$[\mathfrak{I}_\pm, \mathcal{R}_5] = \pm 2\mathfrak{L}_\pm \quad \text{(C15)}$$

give us

$$F_{kl} = 0, \quad \text{(C16)}$$

$$c_{kl} + \epsilon[d_{kl}(m) - d_{kl}(m\pm 1)] = \delta_{kl}[\epsilon(\xi^2-\epsilon^2)^{-\frac{1}{2}} \pm 2t\xi\epsilon^{-1} \mp 2(m\pm 1) \times \epsilon^{-1}(\xi^2-\epsilon^2)^{\frac{1}{2}}], \quad \text{(C17)}$$

from which we get

$$c_{kl} = 0, \quad k, l \text{ not all } 0, \quad \text{(C18)}$$

$$d_{kl}(m) = d_{kl}, \quad k, l \text{ not all } 0, \quad \text{(C19)}$$

$$d_{00}(m) = m^2\epsilon^{-2}(\xi^2-\epsilon^2)^{\frac{1}{2}} - 2tm\xi\epsilon^{-2} + d_{00}, \quad \text{(C20)}$$

where d_{00} is independent of m.

Finally, from

$$[L_{03}, \mathcal{R}_5] = i\mathcal{R}_5 \quad \text{(C21)}$$

and the unitarity condition, we get

$$a_{00} = \eta(\xi^2-\epsilon^2)^{-\frac{1}{2}} - \eta^{-1}(\xi^2-\epsilon^2)^{\frac{1}{2}}, \quad \text{(C22)}$$

$$b_{00} = \xi^{-1}[4(\xi^2-\epsilon^2)^{\frac{1}{2}} + \epsilon^2(\xi^2-\epsilon^2)^{-\frac{1}{2}}], \quad \text{(C23)}$$

$$c_{00} = \epsilon(\xi^2-\epsilon^2)^{-\frac{1}{2}} - \epsilon^{-1}(\xi^2-\epsilon^2)^{\frac{1}{2}}, \quad \text{(C24)}$$

$$a_{0\pm}(s,t) = \mp\epsilon\xi^{-2}(\xi^2+\eta^2)^{\frac{1}{2}}[(s\mp t)(s\pm t+1)]^{\frac{1}{2}}, \quad \text{(C25)}$$

$$a_{\pm l} = 0, \quad \text{(C26)}$$

$$b_{kl} = c_{kl} = 0, \quad k, l \text{ not all } 0. \quad \text{(C27)}$$

The representation for $\mathcal{R}_5$ can now be written as

$$\begin{aligned}
\mathcal{R}_5 |\eta, \xi, \epsilon; s, t, m\rangle
&\Big[(\xi^2-\epsilon^2)^{\frac{1}{2}}\Big(-\frac{\partial^2}{\partial\eta^2} + \frac{\partial^2}{\partial\xi^2} - \frac{\partial^2}{\partial\epsilon^2} + 2\frac{\eta}{\xi}\frac{\partial}{\partial\eta}\frac{\partial}{\partial\xi}\Big) \\
&+ [\eta(\xi^2-\epsilon^2)^{-\frac{1}{2}} - \eta^{-1}(\xi^2-\epsilon^2)^{\frac{1}{2}}]\frac{\partial}{\partial\eta} \\
&+ \xi^{-1}[4(\xi^2-\epsilon^2)^{\frac{1}{2}} + \epsilon^2(\xi^2-\epsilon^2)^{-\frac{1}{2}}]\frac{\partial}{\partial\xi} \\
&+ [\epsilon(\xi^2-\epsilon^2)^{-\frac{1}{2}} - \epsilon^{-1}(\xi^2-\epsilon^2)^{\frac{1}{2}}]\frac{\partial}{\partial\epsilon} \\
&+ m^2\epsilon^{-2}(\xi^2-\epsilon^2)^{\frac{1}{2}} - 2tm\xi\epsilon^{-2} + d_{00}\Big] \\
&\times |\eta, \xi, \epsilon; s, t, m\rangle \\
&+ \Big(-\epsilon\xi^{-2}(\xi^2+\eta^2)^{\frac{1}{2}}[(s-t)(s+t+1)]^{\frac{1}{2}}\frac{\partial}{\partial\eta} \\
&\qquad + d_{0+}(m)\Big)|\eta, \xi, \epsilon; s, t+1, m\rangle \\
&+ \Big(\epsilon\xi^{-2}(\xi^2+\eta^2)^{\frac{1}{2}}[(s+t)(s-t+1)]^{\frac{1}{2}}\frac{\partial}{\partial\eta} \\
&\qquad + d_{0-}(m)\Big)|\eta, \xi, \epsilon; s, t-1, m\rangle \\
&+ \sum_{l=-}^{+} d_{+l}(m)\,|\eta, \xi, \epsilon; s+1, t+l, m\rangle \\
&+ \sum_{l=-}^{+} d_{-l}(m)\,|\eta, \xi, \epsilon; s-1, t+l, m\rangle, \quad \text{(C28)}
\end{aligned}$$

and the unitarity condition gives us the following relations:

$$d_{++}(s, t) = d_{--}(s + 1, t + 1), \tag{C29}$$

$$d_{+-}(s, t) = d_{-+}(s + 1, t - 1), \tag{C30}$$

$$d_{+0}(s, t) = d_{-0}(s + 1, t). \tag{C31}$$

We next consider

$$[\mathfrak{L}_{\pm 6}, \mathfrak{R}_5] = 0, \tag{C32}$$

which, after some lengthy algebra, gives us

$$d_{++}(s, t, m) = [(s + t + 1)(s + t + 2)]^{\frac{1}{2}}\epsilon\xi^{-1}C(\eta; \xi^2 - \epsilon^2), \tag{C33}$$

$$d_{+0}(s, t, m) = 2[(s + t + 1)(s - t + 1)]^{\frac{1}{2}}\xi^{-1}\eta^{-1}(\xi^2 + \eta^2)^{\frac{1}{2}} \times (\xi^2 - \epsilon^2)^{\frac{1}{2}}C(\eta; \xi^2 - \epsilon^2), \tag{C34}$$

$$d_{+-}(s, t, m) = -[(s - t + 1)(s - t + 2)]^{\frac{1}{2}}\epsilon\zeta^{-1}C(\eta; \xi^2 - \epsilon^2), \tag{C35}$$

where $C(\eta; \xi^2 - \epsilon^2)$ is a function of η and $\xi^2 - \epsilon^2$ only and

$$d_{0\pm}(s, t, m) = [(s \mp t)(s \pm t + 1)]^{\frac{1}{2}} \times \{\mp\epsilon\xi^{-2}(\xi^2 + \eta^2)^{\frac{1}{2}}[\pm t\eta^{-1} + \eta^{-1} + \tfrac{1}{2}\eta(\xi^2 + \eta^2)^{-1}] + \epsilon\xi^{-1}B(\eta, \xi^2 - \epsilon^2)\}. \tag{C36}$$

From

$$[\mathfrak{R}_5, \mathfrak{L}_\pm] = \pm\mathfrak{R}_\pm, \tag{C37}$$

we obtain representations for $\mathfrak{R}_\pm$, while the commutation relations

$$[\mathfrak{L}_+, \mathfrak{R}_+] = 0, \tag{C38}$$

$$[\mathfrak{R}_+, \mathfrak{L}_-] = 2\mathfrak{R}_5 \tag{C39}$$

further restrict the functions d_{00}, $B(\eta, \xi^2 - \epsilon^2)$, and $C(\eta, \xi^2 - \epsilon^2)$. The final complete results for $\mathfrak{R}_\mu$ are given in Eqs. (4.11)–(4.14), where we have also used

$$i[\mathfrak{R}_5, \mathfrak{L}_{56}] = \mathfrak{R}_6 \tag{C40}$$

and

$$[\mathfrak{L}_+, \mathfrak{R}_6] = 0. \tag{C41}$$

All other commutation relations can now be shown to be satisfied automatically.

* This work was supported in part by the U.S. Atomic Energy Commission.

† A summary of this work was presented at the Symposium on de Sitter and Conformal Groups, Boulder, Colorado, June 29–July 3, 1970.

[1] See, for instance, P. Carruthers and M. Gell-Mann (unpublished) and C. J. Isham, A. Salam, and J. Strathdee, ICTP, Trieste, Preprint IC/70/17, 1970.

[2] See, for instance, A. O. Barut, *Springer Tracts on Modern Physics: Ergebnisse der Exakten Naturwissenschaften* (Springer-Verlag, Berlin, 1969), Vol. 50, and T. Yao (to be published).

[3] K. Johnson (private communication).

[4] See, for instance, W. Miller, Jr., *Lie Theory and Special Functions* (Academic, New York, 1968); J. D. Talman, *Special Functions, A Group Theoretic Approach* (Benjamin, New York, 1968); N. Mukunda, J. Math. Phys. **10**, 2086 (1969), and S. J. Chang and L. O'Raifeartaigh, *ibid.*, 21 (1969).

[5] T. Yao, J. Math. Phys. **8**, 1931 (1967) (Paper I); **9**, 1615 (1968) (Paper II). [Other references on $SU(2, 2)$ can be found here.]

[6] V. Bargmann, Ann. Math. **48**, 568 (1947); A. O. Barut and C. Fronsdal, Proc. Roy. Soc. (London) **A287**, 532 (1965).

[7] Compare with W. H. Klink, J. Math. Phys. **10**, 606 (1969), and G. Mack and I. Todorov, *ibid.*, 2078 (1969).

[8] M. A. Naimark, *Linear Representations of the Lorentz Group* (Pergamon, New York, 1964), pp. 116–19.

[9] H. Buchholz, *The Confluent Hypergeometric Function* (Springer, New York, 1969), p. 85.

[10] *Note added in proof:* There is an omission in the discussion of the most degenerate principal continuous series and the most degenerate complementary continuous series. In both cases we may ask the question of the reduction of these series with respect to the "spacelike" representations of $E(3, 1)$. Parallel to the derivation of Eqs. (7.71)–(7.87), we define

$$\tilde{f}(\omega, \xi, \epsilon, \lambda) \equiv \langle j = 0, \mu = 0; k = 0, \nu = 0; \lambda \mid \omega, \xi, \epsilon; r = 1, t = 0, m = 0\rangle,$$

where the trivial 1-dimensional UIR of $SU(1, 1)$ is used, and obtain

$$\left[x\left(\frac{\partial^2}{\partial x^2} + \frac{1}{x}\frac{\partial}{\partial x} - \frac{1}{4} + \frac{1}{4}\frac{\rho^2}{x^2}\right) + y\left(\frac{\partial^2}{\partial y^2} + \frac{1}{y}\frac{\partial}{\partial y} - \frac{1}{4} + \frac{1}{4}\frac{\rho^2}{y^2}\right)\right] \times \tilde{h}(x, y, \lambda) = 0, \tag{7.81'}$$

$$\left[x^2\left(\frac{\partial^2}{\partial x^2} + \frac{2}{x}\frac{\partial}{\partial x} - \frac{1}{4} + \frac{1}{2}\frac{\lambda}{x}\right) - y^2\left(\frac{\partial^2}{\partial y^2} + \frac{1}{y}\frac{\partial}{\partial y} - \frac{1}{4} - \frac{1}{2}\frac{\lambda}{y}\right)\right] \times \tilde{h}(x, y, \lambda) = 0, \tag{7.82'}$$

with

$$x = \xi + (\xi^2 - \omega^2)^{\frac{1}{2}}, \tag{7.80a'}$$

$$y = \xi - (\xi^2 - \omega^2)^{\frac{1}{2}}, \tag{7.80b'}$$

and the final result

$$\tilde{f}(\omega, \xi, \epsilon; \lambda) = \tilde{c}(\lambda, \rho)(\xi^2 - \epsilon^2)^{-\frac{1}{4}}(\xi^2 - \omega^2)^{-\frac{1}{4}}\xi^{-1} \times \{W_{\frac{1}{2}(\lambda+1), \frac{1}{2}i\rho}(x)W_{-\frac{1}{2}(\lambda-1), \frac{1}{2}i\rho}(y) + \tfrac{1}{4}(\lambda^2 + \rho^2)W_{\frac{1}{2}(\lambda-1), \frac{1}{2}i\rho}(x)W_{-\frac{1}{2}(\lambda+1), \frac{1}{2}i\rho}(y)\} \tag{7.87'}$$

The analogous result [Eq. (7.89′)] may be obtained by substituting $i\rho \to \sigma$ in Eq. (7.87′). Therefore, the above two series contain both "timelike" and "spacelike" representations of $E(3, 1)$.

I wish to thank Dr. N. Macfadyen for correspondence and a preprint of his paper, "Conformal Group in a Poincaré Basis. I. Principal Continuous Series" [J. Math. Phys., (to be published) (1971)], which led to the discovery of the oversight in my manuscript.

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 12, NUMBER 3 MARCH 1971

Periodic Orbits and Classical Quantization Conditions

MARTIN C. GUTZWILLER*
IBM Watson Laboratory, Columbia University, New York, New York 10025

(Received 27 August 1970)

The relation between the solutions of the time-independent Schrödinger equation and the periodic orbits of the corresponding classical system is examined in the case where neither can be found by the separation of variables. If the quasiclassical approximation for the Green's function is integrated over the coordinates, a response function for the system is obtained which depends only on the energy and whose singularities give the approximate eigenvalues of the energy. This response function is written as a sum over all periodic orbits where each term has a phase factor containing the action integral and the number of conjugate points, as well as an amplitude factor containing the period and the stability exponent of the orbit. In terms of the approximate density of states per unit interval of energy, each stable periodic orbit is shown to yield a series of δ functions whose locations are given by a simple quantum condition: The action integral differs from an integer multiple of h by half the stability angle times $\hbar$. Unstable periodic orbits give a series of broadened peaks whose half-width equals the stability exponent times $\hbar$, whereas the location of the maxima is given again by a simple quantum condition. These results are applied to the anisotropic Kepler problem, i.e., an electron with an anisotropic mass tensor moving in a (spherically symmetric) Coulomb field. A class of simply closed, periodic orbits is found by a Fourier expansion method as in Hill's theory of the moon. They are shown to yield a well-defined set of levels, whose energy is compared with recent variational calculations of Faulkner on shallow bound states of donor impurities in semiconductors. The agreement is good for silicon, but only fair for the more anisotropic germanium.

INTRODUCTION

The Green's function of a simple quantum mechanical system can be approximated with the help of quantities which arise from the corresponding classical mechanical system. This idea has been explored in a series of three papers,[1] where it was shown that all the well-known results of the WKB method for bound states can be obtained in this manner without any further assumptions or rules. In the last paper the concept of a semiclassical density of states per unit interval of energy was introduced and tested for a particle in a box and in a spherically symmetric potential. It turned out that the classical periodic orbits alone determine this density of states and are, therefore, responsible for the occurrence of stationary states, i.e., regions of large density along the energy axis.

The present paper extends this last idea to particles in a potential where neither Schrödinger's equation nor the corresponding classical equations of motion can be separated, i.e., reduced to as many independent motions as degrees of freedom. In other words, the phase-integral approximation is extended to systems which are not multiply periodic, and quantization conditions *à la* Bohr and Sommerfeld are derived for cases in which none had been formulated so far. The arguments and the results involve only notions from classical mechanics which were well known to the astronomers at the end of the 19th century, in particular, Hill and Poincaré. Unfortunately, the present day education in physics ignores these classical results. It seemed, therefore, necessary to explain them in some detail in the context of the present work.

The first section recalls the quasiclassical approximation for the Green's function and its reduction to a density of states. Then, we show that only periodic orbits contribute to the latter quantity if we restrict ourselves to the leading term of an expansion in powers of Planck's quantum. The density of states appears as a sum over all periodic orbits for a given energy. Each term in this sum is an integral over the periodic orbit which has to be evaluated. This task is carried out in Secs. 2, 3, and 4. Great care is taken to stay away from any special assumptions about the mechanical system except that it has a Hamiltonian which is independent of time. The original expressions involve only the action integral between two points along a classical trajectory. But, after some rather elementary calculations, everything can be expressed in terms of the action around the periodic orbits as a function of the energy, the number of conjugate points, and the stability exponents.

The result is most conveniently written as a kind of response function for the system in a complex energy plane such that its discontinuity along the real energy axis equals the density of states. This response function is remarkably similar to the so-called zeta function which mathematicians have investigated in order to survey and classify the periodic orbits of abstract mechanical systems.[2] Actually, we are tempted to propose our response function as the appropriate object of study rather than the zeta function, because the latter does not include any information about conjugate points and stability in its definition. The behavior of trajectories in the neighborhood of a

periodic orbit is, of course, a very important ingredient if we think of Feynman's approach to quantum mechanics. But, if the periodic orbits are the key to the understanding of a mechanical system as Poincaré suggested, we will surely have to see them as tied in with their environment of neighboring trajectories; and this environment is most aptly described by the number of conjugate points and the stability exponent, independently of any implications for quantum mechanics.

The formulas at the end of Sec. 4 are a straightforward consequence of the general expression for the approximate Green's function which has been used ever since it was derived in the first paper of this series. These formulas have a simple physical interpretation, but they are unsatisfactory in one respect. Every time a particular periodic orbit has a conjugate point after nearly one period, its contribution to the response function is very large. This feature is not surprising because the wave amplitude at a focus is infinite according to ray optics, although it remains finite according to wave optics. The large amplitude near a conjugate point is an artifact of our approximation, and we have modified the formulas to get rid of this effect in agreement with the known behavior of a wavefield near a focal point. This modification, however, looks like an ad hoc improvement, contrary to the treatment of focal points in the preceding paper. The situation is discussed in Sec. 5 for stable orbits and in Sec. 6 for unstable ones. In the first case we end up with a series δ-function spikes in the density of states which are located by a simple quantization condition involving explicitly the stability angle. In the second case we get broadened peaks whose width is given by the instability exponent and whose maxima are given by an ordinary quantization condition such as are found in the earlier papers. The most general case of a periodic orbit in three dimensions with a complex exponent (rather than a purely imaginary or purely real one) is not discussed, but it is expected to combine the features of the two simpler cases.

The last two sections of this paper treat a simple example of a nonseparable problem of some practical importance, the bound states of an electron near a donor impurity of a semiconductor. The Hamiltonian differs from that for the hydrogen atom in having an anisotropic mass tensor, although the potential is due to an ordinary (spherically symmetric) Coulomb field. The virial theorem is still true, however, and a simplified form of Kepler's third law is valid. But the periodic orbits are isolated at any given energy, and have to be found by numerical methods unless the anisotropy is very small.

Only one type of periodic orbit is investigated because it can be obtained by the same kind of Fourier expansion which Hill used in his classic theory of the motion of the moon. This orbit corresponds to a state of maximum angular momentum compatible with the principal quantum number. It is always unstable and has the same number of conjugate points as the orbits of the ordinary Kepler problem. The formula for these energy levels looks, therefore, almost like the Balmer formula except for the normalization of the energy which depends on the anisotropy. A comparison of these semiclassical, approximate energies is made with the variational calculations of Kohn and Luttinger[3] and, more recently, of Faulkner.[4] It turns out that Faulkner's designations of levels have to be changed before they can be made to correspond to our results. Although the agreement is good for silicon and reasonable for germanium, a final assessment of our quasiclassical approach has to wait until other types of periodic orbits have been evaluated and until there is a complete enumeration of the possible energies and their symmetries as far as classical mechanics can tell.

In this way we are finally faced with the following crucial question: What is the relation between the periodic orbits in the classical system and the energy levels of the corresponding quantum system? It seems that for the low-lying states there is a one-to-one correspondence if the classical orbits are quantized according to the rules of this paper. On the other hand, it also looks as if the more complicated orbits are less stable and their peaks in the density of states, therefore, wider, i.e., less distinct. Even more serious is the fact that there is usually more than a countable number of periodic orbits in a mechanical system,[5] whereas the bound states of an ordinary Hamiltonian are countable. It is not clear at all at what point the apparent simple relation between periodic orbits and discrete energy levels breaks down, particularly since the semiclassical approach to quantum mechanics is supposed to be the better the larger the quantum number.

1. THE RESPONSE FUNCTION

The Hamiltonian $H(pq)$ of the system is assumed to allow only for a discrete group of symmetry operations. Both spherical and cylindrical symmetry are explicitly excluded from the following investigation because they present special problems which have been examined in an earlier paper. The anisotropy of the Hamiltonian can arise in the kinetic energy as well as in the potential energy. Therefore, it will not be assumed that the velocity $\dot{q}$ is simply proportional to the momentum p.

The two quantities will be carefully separated whenever they arise. On the other hand, the mass will not appear at all except in the discussion of special examples.

The starting point is the quasiclassical approximation $\tilde{G}(q''q'E)$ to the Green's function $G(q''q'E)$ for the position coordinates q' and q'' and the energy E. According to the preceding papers,

$$\tilde{G}(q''q'E) = -\frac{1}{2\pi\hbar^2} \sum_{\substack{\text{classical} \\ \text{trajectories}}} (|D_s|)^{\frac{1}{2}} \times \exp\left(\frac{S(q''q'E)}{\hbar} - \text{phases}\right), \quad (1)$$

$$S(q''q'E) = \int_{q'}^{q''} p\, dq, \quad \text{along the classical trajectory}, \quad (2)$$

and D_s is the 4×4 determinant of the mixed second derivatives of $S(q''q'E)$,

$$D_s = \begin{vmatrix} \frac{\partial^2 S}{\partial q' \partial q''} & \frac{\partial^2 S}{\partial q' \partial E} \\ \frac{\partial^2 S}{\partial E \partial q''} & 0 \end{vmatrix}. \quad (3)$$

The "phases" are given by $\pi/2$ times the number of conjugate points along the trajectory. The above formulas are written for three dimensions, but they are valid for two dimensions provided that we insert an additional factor $(2\pi\hbar)^{\frac{1}{2}} \exp(i\pi/4)$.

As in the preceding paper, we integrate over all values $q' = q''$ to get the response function

$$g(E) = \int d^3q\, G(qqE) = \sum_j \frac{1}{E - E_j}, \quad (4)$$

where j labels the eigenstates of the quantum mechanical system and E_j the corresponding energies. The quasiclassical approximation $\tilde{g}(E)$ requires that $\tilde{G}(qqE)$ be integrated, and we assume that this can be done separately for each term in the summation over all classical trajectories.

At this point, a simple, but essential, argument enters into the discussion. If we pick a particular classical trajectory whose initial point q' coincides with the final point q'', this trajectory exists even when the point $q' = q'' = q$ is varied over a certain neighborhood. In keeping with the general methods of approximating quantum mechanics by classical mechanics, we assume that the variation with q of the phase factor $\exp(iS/\hbar)$ is fast compared to the variation of the amplitude factor $(|D_s|)^{\frac{1}{2}}$. The rate of variation is determined by the partial derivative

$$\frac{\partial S(qqE)}{\partial q} = \left(\frac{\partial S(q''q'E)}{\partial q'} + \frac{\partial S(q''q'E)}{\partial q''}\right)_{q'=q''=q} = p'' - p', \quad (5)$$

where p' and p'' are the initial and final momenta of the classical trajectory which starts at q and ends at q. The integral over q becomes very small because of destructive interference unless $p'' = p'$, i.e., the classical trajectory is periodic. Therefore, in discussing $\tilde{g}(E)$, the summation over the classical trajectories can be restricted to a summation over the periodic orbits.

The integration over $q' = q'' = q$ is most conveniently performed by introducing a coordinate system where q_1 varies along the periodic orbit and the remaining q_2 and q_3 are perpendicular to the periodic orbit. For a given point $\bar{q}$ on the periodic orbit, we expand

$$S(qqE) = S(\bar{q}\bar{q}E) + \left(\frac{\partial S}{\partial q'} + \frac{\partial S}{\partial q''}\right)_{q'=q''=\bar{q}} \cdot \delta q + \frac{1}{2}\left(\frac{\partial^2 S}{\partial q' \partial q'} + 2\frac{\partial^2 S}{\partial q' \partial q''} + \frac{\partial^2 S}{\partial q'' \partial q''}\right)_{q'=q''=\bar{q}} \cdot \delta q\, \delta q, \quad (6)$$

where δq has the two components $q_2 - \bar{q}_2$ and $q_3 - \bar{q}_3$. The linear term vanishes everywhere along the periodic orbit, and we assume that the quadratic term is singular only for isolated values of $\bar{q}$.

The integration over q_2 and q_3 is performed by the stationary phase method for each value of $\bar{q}$ along the periodic orbit. The result is straightforward,

$$\tilde{g}(E) = -\frac{1}{\hbar} \sum_{\substack{\text{periodic} \\ \text{orbits}}} \oint d\bar{q} (|D_s|)^{\frac{1}{2}} \times \exp i\left(\frac{S(E)}{\hbar} - \text{phases}\right) \exp(\pm\tfrac{1}{4}i\pi \pm \tfrac{1}{4}i\pi) \Big/ \left(\det\left|\frac{\partial^2 S}{\partial q' \partial q'} + 2\frac{\partial^2 S}{\partial q' \partial q''} + \frac{\partial^2 S}{\partial q'' \partial q''}\right|_{2,3}\right)^{\frac{1}{2}}, \quad (7)$$

where the double signs refer to the signs of the eigenvalues for the quadratic term in (6) and the index 2, 3 on the last determinant refers to the variables perpendicular to the periodic orbit at any point $\bar{q}$. The action integral $S(\bar{q}\bar{q}E)$ has been written as $S(E)$ because it is obviously independent of the point $\bar{q}$.

The 2-dimensional case differs from the above only in having $\exp(\pm\frac{1}{4}i\pi \pm \frac{1}{4}i\pi)$ replaced by $\exp(\frac{1}{4}i\pi \pm \frac{1}{4}i\pi)$ and the determinant in the denominator consisting of the second derivatives with respect to q_2 only.

The integral over $\bar{q}$ has to be evaluated in terms of simple quantities which belong to the periodic orbit as a whole rather than to any particular point of it. This will be done in the next sections.

2. SPECIAL COORDINATES FOR A PERIODIC ORBIT

The derivatives of $S(q'q''E)$ will always be evaluated on the periodic orbits, i.e., for $q' = q'' = \bar{q}$, in the Secs. 2, 3, and 4. But this fact will not be specifically mentioned. Also we shall always use a coordinate system where q_1 is parallel to the periodic orbit.

The partial differential equation for $S(q''q'E)$ will be used in both forms

$$H\left(-\frac{\partial S}{\partial q'}, q'\right) = E, \tag{8}$$

$$H\left(\frac{\partial S}{\partial q''}, q''\right) = E. \tag{9}$$

From their derivatives with respect to E, we get

$$-\frac{\partial H}{\partial p'}\cdot\frac{\partial^2 S}{\partial E \partial q'} = 1,$$
$$\frac{\partial H}{\partial p''}\cdot\frac{\partial^2 S}{\partial E \partial q''} = 1. \tag{10}$$

Since $\dot{q} = \partial H/\partial p = (|\dot{q}|, 0, 0)$ in our special coordinate system, it follows that

$$\frac{\partial^2 S}{\partial E \partial q_1'} = -\frac{1}{|\dot{q}|}, \quad \frac{\partial^2 S}{\partial E \partial q_1''} = \frac{1}{|\dot{q}|}. \tag{11}$$

Furthermore, by differentiating (8) with respect to q_i'' and (9) with respect to q_i', we get

$$-\sum_j \frac{\partial H}{\partial p_j'}\cdot\frac{\partial^2 S}{\partial q_i'' \partial q_j'} = 0,$$
$$\sum_j \frac{\partial H}{\partial p_j''}\cdot\frac{\partial^2 S}{\partial q_i' \partial q_j''} = 0. \tag{12}$$

Again, with $\dot{q} = \partial H/\partial p = (|\dot{q}|, 0, 0)$, it follows that

$$\frac{\partial^2 S}{\partial q_i'' \partial q_1'} = \frac{\partial^2 S}{\partial q_i' \partial q_1''} = 0. \tag{13}$$

The determinant D_s is thereby reduced to

$$D_s = \frac{1}{|\dot{q}|^2}\begin{vmatrix} \frac{\partial^2 S}{\partial q_2' \partial q_2''} & \frac{\partial^2 S}{\partial q_2' \partial q_3''} \\ \frac{\partial^2 S}{\partial q_3' \partial q_2''} & \frac{\partial^2 S}{\partial q_3' \partial q_3''} \end{vmatrix}. \tag{14}$$

The first factor in (14) combines with dq in (7) to give $dt = dq/\dot{q}$. If the remaining factors in (7) remain constant along the periodic orbit, the integral over q yields $\int dt = T$, the period, i.e., the time required for the particle to go around the periodic orbit.

In order to see that the remaining terms in (7) are, indeed, constant along the periodic orbit, we examine the two related concepts of monodromy matrix and area preserving maps for a periodic orbit. They are explained in some of the standard texts on classical mechanics.[6] Our task is to express them in terms of the second derivatives of $S(q''q'E)$.

3. TRAJECTORIES CLOSE TO THE PERIODIC ORBIT

Consider the trajectory which starts with the coordinates $q' = q + \xi'$ and the momenta $p' = p + \eta'$ in the neighborhood of the coordinates q and momenta p on the periodic orbit. After a time T equal to the period of the periodic orbit, the particle will be at the coordinates $q'' = q + \xi''$ and the momenta $p'' = p + \eta''$. The 6×6 matrix which is defined by

$$\begin{pmatrix} \xi'' \\ \eta'' \end{pmatrix} = M \begin{pmatrix} \xi' \\ \eta' \end{pmatrix} \tag{15}$$

is called the monodromy matrix.

On the other hand, consider a trajectory of energy E which starts with a coordinate $q_1' = q_1$. After a time t which is close to T, this trajectory will reach an end point where again $q_1'' = q_1$. Thus, the 4-dimensional subspace of phase space where $q_1 = \text{const}$ and $E = \text{const}$ is mapped into itself. This mapping preserves the 4-dimensional volume and has a fix point at the periodic orbit. In the neighborhood of the periodic orbit the mapping is given by a linear transformation θ. Our goal is to express both θ and M in terms of the second derivatives of $S(q''q'E)$ at the periodic orbit.

M contains some information which is of no great physical interest. If $\xi' = \dot{q}\delta t$ and $\eta' = \dot{p}\delta t$ with some arbitrary small δt, one finds that $\xi'' = \dot{q}\delta t$ and $\xi'' = \dot{p}\delta t$ because the trajectory coincides with the periodic orbit. Moreover, there is a periodic orbit whose energy differs by some arbitrary small δE from the original periodic orbit. Its initial coordinates and momenta are a little harder to find.

Starting from the general formulas

$$p' = -\frac{\partial S(q'q''E)}{\partial q'}, \quad p'' = \frac{\partial S(q'q''E)}{\partial q''}, \tag{16}$$

one finds at once that

$$\eta' = -\frac{\partial^2 S}{\partial q' \partial q'}\xi' - \frac{\partial^2 S}{\partial q' \partial q''}\xi'' - \frac{\partial^2 S}{\partial q' \partial E}\delta E,$$
$$\eta'' = \frac{\partial^2 S}{\partial q'' \partial q'}\xi' + \frac{\partial^2 S}{\partial q'' \partial q''}\xi'' + \frac{\partial^2 S}{\partial q'' \partial E}\delta E. \tag{17}$$

The conditions $\eta' = \eta'' = \eta$ and $\xi' = \xi'' = \xi$ for a periodic orbit give the equations

$$\sum_j \left(\frac{\partial^2 S}{\partial q_i' \partial q_j'} + \frac{\partial^2 S}{\partial q_i' \partial q_j''} + \frac{\partial^2 S}{\partial q_i'' \partial q_j'} + \frac{\partial^2 S}{\partial q_i'' \partial q_j''}\right)\xi_j + \left(\frac{\partial^2 S}{\partial q_i' \partial E} + \frac{\partial^2 S}{\partial q_i'' \partial E}\right)\delta E = 0. \quad (18)$$

Neither the first term in (18) nor the second has any components in the direction q_1. For the second term this follows directly from (11), whereas for the first term we can use (12) together with

$$-\sum_j \frac{\partial H}{\partial p_j'}\frac{\partial^2 S}{\partial q_i' \partial q_j'} + \frac{\partial H}{\partial q_i'} = 0,$$
$$+\sum_j \frac{\partial H}{\partial p_j''}\frac{\partial^2 S}{\partial q_i'' \partial q_j''} + \frac{\partial H}{\partial q_i''} = 0, \quad (19)$$

and $\dot{q} = \partial H/\partial p = (|\dot{q}|, 0, 0)$. It is equally obvious that (18) cannot be solved for $\delta E = 0$ with a vector ξ, where second and third components vanish. On the other hand, Eqs. (18) do have a unique solution provided that the quadratic form in (6) is not degenerate or, equivalently, the determinant in the denominator of (7) does not vanish. Thus we find a periodic orbit starting and ending at $\xi_1 = 0$, $\xi_2 = \alpha_2 \delta E$, $\xi_3 = \alpha_3 \delta E$, $\eta_1 = \beta_1 \delta E$, $\eta_2 = \beta_2 \delta E$, and $\eta_3 = \beta_3 \delta E$, where α_2 and α_3 follow from (18), with β_1, β_2, and β_3 then given by (17).

The trajectories which enter into the area preserving map θ are characterized by the conditions $\xi_1' = 0$ and $\delta H = 0$. They can be defined in terms of their intersection with the space $q_1 = \text{const}$, $H = \text{const}$ by four quantities δq_2, δq_3, δp_2, and δp_3 which give the coordinates and momenta in the space $q_1 = \text{const}$, $H = \text{const}$. Thus, instead of ξ and η, we use δt, δE, δq_2, δq_3, δp_2, and δp_3 to determine initial and final points of trajectories in the neighborhood of the periodic orbit. The matrix relating the two reference systems is given by

$$\begin{pmatrix} \xi_1 \\ \eta_1 \\ \xi_2 \\ \xi_3 \\ \eta_2 \\ \eta_3 \end{pmatrix} = \begin{pmatrix} |\dot{q}| & 0 & 0 & 0 & 0 & 0 \\ \dot{p}_1 & \beta_1 & \gamma_2 & \gamma_3 & 0 & 0 \\ 0 & \alpha_2 & 1 & 0 & 0 & 0 \\ 0 & \alpha_3 & 0 & 1 & 0 & 0 \\ \dot{p}_2 & \beta_2 & 0 & 0 & 1 & 0 \\ \dot{p}_3 & \beta_3 & 0 & 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} \delta t \\ \delta E \\ \delta q_2 \\ \delta q_3 \\ \delta p_2 \\ \delta p_3 \end{pmatrix}, \quad (20)$$

where γ_2 and γ_3 are chosen to meet the requirement $\delta H = 0$, which leads to

$$|\dot{q}|\,\gamma_2 + \frac{\partial H}{\partial q_2} = 0, \quad |\dot{q}|\,\gamma_3 + \frac{\partial H}{\partial q_3} = 0. \quad (21)$$

The determinant of the matrix L in (20) equals 1. But L does not constitute a canonical transformation. If J is the 6×6 matrix which defines the bilinear form $\xi_1''\eta_1' - \eta_1''\xi_1' + \xi_2''\eta_2' - \eta_2''\xi_2' + \xi_3''\eta_3' - \eta_3''\xi_3'$, the matrix

$$L^+JL = \begin{pmatrix} 0 & 1 & 0 & 0 & 0 & 0 \\ -1 & 0 & -\beta_2 & -\beta_3 & \alpha_2 & \alpha_3 \\ 0 & \beta_2 & 0 & 0 & 1 & 0 \\ 0 & \beta_3 & 0 & 0 & 0 & 1 \\ 0 & -\alpha_2 & -1 & 0 & 0 & 0 \\ 0 & -\alpha_2 & 0 & -1 & 0 & 0 \end{pmatrix} \quad (22)$$

differs from J by the occurrence of α's and β's in the second row and second column. (L^+ is the Hermitian conjugate of L, i.e., the transpose of L in this case.) This defect does not impair any of the following arguments, however.

The monodromy matrix is given in the new coordinate system by

$$L^{-1}ML = \begin{pmatrix} 1 & \otimes & \otimes & \otimes & \otimes & \otimes \\ 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & & & & \\ 0 & 0 & & \text{reduced matrix} & & \\ 0 & 0 & & \text{for trajectories} & & \\ 0 & 0 & & \text{with } \delta H = 0 & & \\ 0 & 0 & & \text{and } \delta q_1 = 0 & & \end{pmatrix} \quad (23)$$

The nonzero elements in the first row are due to the fact that all trajectories run through a constant time interval T in the monodromy matrix. Their end points are not in the subspace $\delta q_1 = 0$. If the running time of the trajectories with $\delta t = 0$ and $\delta E = 0$ is shortened by the amount calculated from the first row, their end points will be in the subspace $\delta q_1 = 0$. Therefore, the 4×4 matrix in the lower right-hand corner of (23) is exactly the area preserving map θ.

4. THE STABILITY INDICES OF THE PERIODIC ORBIT

The area preserving transformation θ can be obtained from the general formulas (16):

$$\delta p_i' = -\sum_{j=2}^{3} \frac{\partial^2 S}{\partial q_i' \partial q_j'} \delta q_j' - \sum_{j=2}^{3} \frac{\partial^2 S}{\partial q_i' \partial q_j''} \delta q_j'',$$

$$\delta p_i'' = \sum_{j=2}^{3} \frac{\partial^2 S}{\partial q_i'' \partial q_j'} \delta q_j' + \sum_{j=2}^{3} \frac{\partial^2 S}{\partial q_i'' \partial q_j''} \delta q_j'', \qquad (24)$$

where the index i assumes only the values 2 and 3. In terms of the corresponding 2×2 matrices

$$a = \frac{\partial^2 S}{\partial q' \partial q'}, \quad b = \frac{\partial^2 S}{\partial q' \partial q''}, \quad c = \frac{\partial^2 S}{\partial q'' \partial q''}, \qquad (25)$$

Eqs. (24) can be written more simply as

$$\delta p' = -a\delta q' - b\delta q'',$$
$$\delta p'' = b^{+}\delta q' + c\delta q''. \qquad (26)$$

The area preserving transformation θ can be written as

$$\delta q'' = A\delta q' + B\delta p',$$
$$\delta p'' = C\delta q' + D\delta p'. \qquad (27)$$

A comparison with (26) yields immediately

$$A = -b^{-1}a, \qquad B = -b^{-1},$$
$$C = b^{+} - cb^{-1}a, \qquad D = -cb^{-1}. \qquad (28)$$

Only the existence of the 2×2 matrix b^{-1} is required; the determinant of b occurs in D_s according to (14), so that when b is singular, the amplitude in (7) vanishes.

The eigenvalues of θ are given by the roots of the characteristic determinant

$$F(\lambda) = \det \begin{vmatrix} A - \lambda I & B \\ C & D - \lambda I \end{vmatrix}. \qquad (29)$$

After the expressions (28) have been inserted into (29), we add $-c$ times the first line to the second line. In this manner, we find the following (progressively simpler) formulas:

$$F(\lambda) = \begin{vmatrix} -b^{-1} - \lambda I & -b^{-1} \\ b^{+} + \lambda c & -\lambda I \end{vmatrix}$$
$$= \frac{1}{|b|} \begin{vmatrix} -a - \lambda b & -I \\ b^{+} + \lambda c & -\lambda I \end{vmatrix}$$
$$= \frac{1}{|b|} \begin{vmatrix} -a - \lambda b & -I \\ b^{+} + \lambda a + \lambda c + \lambda^2 b & 0 \end{vmatrix}$$
$$F(\lambda) = \frac{\det |b^{+} + \lambda a + \lambda c + \lambda^2 b|}{\det |b|}. \qquad (30)$$

The characteristic determinant has been reduced from 4×4 to 2×2. Upon comparing the last expression with the amplitude in (7), it is obvious that the ratio of the two determinants in (7) is exactly equal to $F(1)$.

It is a general theorem of stability theory, and it is immediately obvious from (30) that the eigenvalues of θ come in pairs which are mutually reciprocal. Since the coefficients in $F(\lambda)$ are real, its roots also come in pairs which are complex conjugate. A number of special cases can be distinguished according as the eigenvalues are real or lie on the unit circle of the complex plane.

In two dimensions we have only two cases:

(1) $F(\lambda)$ has two real roots $\lambda = e^{u}$ and e^{-u}; the periodic orbit is unstable, and we find that

$$F(1) = -4(\sinh \tfrac{1}{2}u)^2. \qquad (31)$$

(2) $F(\lambda)$ has two roots on the unit circle $\lambda = e^{iv}$ and e^{-iv}; the periodic orbit is stable, and we find that

$$F(1) = 4(\sin \tfrac{1}{2}v)^2. \qquad (32)$$

In three dimensions we can have any combination of the two preceding cases plus a new case which arises when $F(\lambda)$ has four complex roots $\lambda = e^{u+iv}$, e^{u-iv}, e^{-u-iv}, and e^{-u+iv}, where $u \neq 0$ and v is not a multiple of π. The periodic orbit shows a complex behavior and we can write

$$F(1) = 16(\sinh^2 \tfrac{1}{2}u \cos^2 \tfrac{1}{2}v + \cosh^2 \tfrac{1}{2}u \sin^2 \tfrac{1}{2}v)^2. \qquad (33)$$

This last case is expected to arise whenever the periodic orbit has no simple symmetry property such as lying in a plane of symmetry for the mechanical system.

It is evident that $F(1)$ is the same for all points along a periodic orbit. The integration over $\tilde{q}$ in (7) can be performed as indicated at the end of Sec. 2. Before writing down the final result for $\tilde{g}(E)$, however, it is necessary to discuss the double signs in (7) together with the stability conditions of the periodic orbit.

5. STABLE ORBITS

Systems with two degrees of freedom will now be discussed. If the periodic orbit is stable, we have $F(1) > 0$. The number of conjugate points along the orbit is even or odd according as

$$\frac{\partial^2 S}{\partial q_2' \partial q_2''} < 0 \quad \text{or} \quad > 0, \qquad (34)$$

and, therefore,

$$\frac{\partial^2 S}{\partial q_2' \partial q_2''} + \frac{\partial^2 S}{\partial q_2' \partial q_2''} + \frac{\partial^2 S}{\partial q_2' \partial q_2''} < 0 \quad \text{or} \quad > 0. \qquad (35)$$

The last condition is responsible for a factor $\exp(\frac{1}{4}i\pi \mp \frac{1}{4}i\pi)$ according to (7), whereas the phase loss due to conjugate points accounts for a factor $+1$, $-i$, -1, and $+i$ according as the number of conjugate points equals 0, 1, 2, or 3, modulo 4. Thus, in all cases the amplitude factor for a stable orbit turns out to be real.

Let us now consider a particular stable periodic orbit and obtain its contribution to the response function $\tilde{g}(E)$. Such an orbit is represented by more than one term in the summation over all periodic orbits because it can be traversed once, twice, three times, etc. The number of conjugate points during one, two, three, etc., periods varies somewhat capriciously. But it can easily be related to the stability angle v which enters into the amplitude according to (32).

The trajectories close to the periodic orbit define an area preserving map between the phase spaces $(\delta p_2', \delta q_2')$ and $(\delta p_2'', \delta q_2'')$ which belong to any two points q' and q'' of the periodic orbit. If we fix q' and let q'' run around the periodic orbit starting at q', the angle v of the map from $(\delta p_2', \delta q_2')$ to $(\delta p_2'', \delta q_2'')$ varies smoothly. Each time v equals a multiple of π, a conjugate point arises. If we disregard any local maxima or minima of v as a function of the time along the periodic orbit and if we choose a direction of motion such that v increases, the number of conjugate points equals the number of times π is contained in v.

The angle v was determined only up to a multiple of 2π in the preceding section. We shall now choose v such that it reflects the number of conjugate points along the periodic orbit appropriately, or in any case as it arises when the point q'' reaches the initial point q' after one, two, three, etc., periods. If v_n is the stability angle after n periods, then we have clearly $v_n = n \cdot v_1$. Similarly, if S_n is the action integral after n periods, we have $S_n = n \cdot S_1$. The factor T as obtained from (14) is independent of the number of periods, however, because it comes from an integration over coordinate space, not over time.

After all these preliminary explanations, we can now combine (32) with (7). The contribution of a particular stable periodic orbit to $\tilde{g}(E)$ becomes simply

$$-\frac{T}{\hbar}\sum_{n=1}^{\infty}\frac{1}{2\sin\frac{1}{2}nv}\exp\left(in\frac{S}{\hbar}\right), \tag{36}$$

where we have written v and S instead of v_1 and S_1. It should be noted that there is no trace of conjugate point bookkeeping in (36). The denominator in each term takes care of that. It is instructive to check how this happens. With no conjugate point there is a factor $\exp(\frac{1}{4}i\pi - \frac{1}{4}i\pi)$ and no phase loss, but $v < \pi$. With one conjugate point there is a factor $\exp(\frac{1}{4}i\pi + \frac{1}{4}i\pi)$ and a phase loss of $-\pi/2$, but $\pi < v < 2\pi$. With two conjugate points there is a factor $\exp(\frac{1}{4}i\pi - \frac{1}{4}i\pi)$ and a phase loss of $-\pi$, but $2\pi < v < 3\pi$ so that $\sin(v/2) < 0$. Finally, with three conjugate points there is a factor $\exp(\frac{1}{4}i\pi + \frac{1}{4}i\pi)$ and a phase loss of $-3\pi/2$, but $3\pi < v < 4\pi$ so that $\sin(v/2) < 0$.

If a particular stable periodic orbit is isolated, as was assumed in the discussion of Sec. 3, it can be followed as the energy E changes. Its contribution to $\tilde{g}(E)$ is given by (36) whose imaginary part will yield the density of states as due to this orbit. Before studying the behavior of (36) as E varies, it is important to understand the physical origin of its main features.

Since the argument of the exponential contains Planck's constant $\hbar$ (divided by 2π), a quantity entirely outside classical mechanics, we can consider $S/\hbar$ and v as varying independently, although both are functions of the energy E. The response function $\tilde{g}(E)$ is always used together with an integration over E. Because of $T = \partial S/\partial E$, we have, therefore, the relation $dE(T/\hbar) = dS/\hbar$. The factor $T/\hbar$ transforms the integration over the energy as independent variable into an integration over the action normalized by Planck's constant. Formula (36) describes, apart from the denominator, the interference of waves which run around the stable periodic orbit once, twice, etc.

The denominator is obviously due to the presence of focal points along the orbit. In particular, the amplitude of any term in (36) is large whenever an even conjugate point is near the starting point of the particle. It is well known, however, that the approximation (1) for the Green's function breaks down near the focal point. As was shown in the preceding paper for certain special examples, the amplitude remains finite, and the phase loss is exactly half what it is upon traversing the focal point completely. The singularities in (36) when nv = multiple of 2π are unphysical and have to be replaced by a more reasonable function of v. Although a detailed investigation of the focal points is required to make (36) really useful, a more intuitive approach gives good results.

Formula (36) gives a finite amplitude near an odd conjugate point, i.e., where $nv = (2l + 1)\pi$ with l integer. We can, therefore, argue that (36) is correct for odd conjugate points, and we have only to see to it that the phase loss is continuous. The phase jumps by π near an even conjugate point have to be smoothed out, and the amplitude has to be reduced to what it is for an odd conjugate point. It is clear that passing a conjugate point always brings a loss (not a gain!) of

phase. The accumulated loss is given by $nv/2$ as discussed earlier. In two dimensions one starts with a phase factor $e^{i\pi/4}$ at the beginning of any trajectory. This initial phase of $\pi/4$ is decreased by $\frac{1}{2}m\pi - \frac{1}{4}\pi$ in the mth conjugate point where the extra $-\pi/4$ corresponds to half the loss of $\pi/2$ upon traversing the mth conjugate point. If $m\pi$ is replaced by v, we end up with the formula

$$-\frac{T}{2\hbar}\sum_{n=1}^{\infty}\exp\left(in\frac{S}{\hbar} - in\frac{v}{2} + i\frac{\pi}{2}\right) \tag{37}$$

to replace (36). Formula (37) coincides with (36) when $nv = (2l+1)\pi$ and replaces the discontinuous phase dependence with occasional infinite amplitudes by a linear phase loss with constant amplitude. Although the above reasoning is hardly foolproof, the final formula (37) incorporates the essential physical features of (36) without its obvious shortcomings.

The formal evaluation of (37) is trivial because it can be considered as a geometric series. Thus, we get

$$\begin{aligned}-\frac{iT}{2\hbar}\frac{\exp i(S/\hbar - v/2)}{1-\exp i(S/\hbar - v/2)} &= \frac{T}{4\hbar}\cot\left(\frac{S}{2\hbar} - \frac{v}{4}\right) + i \\ &= \frac{1}{2}\frac{\partial S}{\partial E}\left[\frac{1}{S - \frac{1}{2}v\hbar} + \sum_{-\infty}^{+\infty}\left(\frac{1}{S - \frac{1}{2}v\hbar - 2m\pi\hbar}\right) + i\right],\end{aligned} \tag{38}$$

where we have used the partial fraction expansion of cot. The factor $\frac{1}{2}$ disappears when we take into account that each periodic orbit can be traversed in two directions, "forward" and "backward."

The response function $\tilde{g}(E)$ has a simple pole of residue 1 at an energy which is given by the condition

$$S = (2m\pi + v/2)\hbar \tag{39}$$

for each stable periodic orbit. In other words, the density of states has a δ-function singularity of strength 1 for each stable periodic orbit which satisfies (39). This condition seems to be the generalization of the classical quantization rule of Bohr and Sommerfeld to mechanical systems which are not multiply periodic, i.e., systems which cannot be decoupled into subsystems of one degree of freedom each, with the help of appropriate constants of motion. Although (39) is simple enough, the author has never seen it derived or even mentioned in any textbook or research paper.

6. UNSTABLE ORBITS

Consider now an unstable periodic orbit in a system with two degrees of freedom. According to (31) we have $F(1) < 0$. The number of conjugate points along the orbit is even or odd according as

$$\frac{\partial^2 S}{\partial q_2' \partial q_2''} < 0 \quad \text{or} \quad > 0, \tag{34}$$

and, therefore,

$$\frac{\partial^2 S}{\partial q_2'^2} + 2\frac{\partial^2 S}{\partial q_2' \partial q_2''} + \frac{\partial^2 S}{\partial q_2''^2} > 0 \quad \text{or} \quad < 0. \tag{40}$$

The last condition is responsible for a factor $\exp(\frac{1}{4}i\pi \pm \frac{1}{4}i\pi)$ according to (7), whereas the phase loss due to conjugate points equals $\pi/2$ times the number of conjugate points. Thus, in all cases the amplitude factor for an unstable orbit is purely imaginary.

The contribution of a particular unstable periodic orbit to $\tilde{g}(E)$ contains again more than one term because the orbit can be traversed once, twice, etc. The number of conjugate points is simply proportional to the number of periods. Thus, if there are l conjugate points for a simple period, there will be $n \cdot l$ conjugate points for n periods. Similarly, the action integral will be nS and the instability exponent nu, if S and u are the action integral and instability exponent for a single period.

In analogy to (36), the contribution of a particular unstable periodic orbit to $\tilde{g}(E)$ is given by the sum

$$-\frac{iT}{\hbar}\sum_{n=1}^{\infty}\frac{1}{2\sinh\frac{1}{2}nu}\exp\left[in\left(\frac{S}{\hbar} - l\frac{\pi}{2}\right)\right]. \tag{41}$$

The bookkeeping of conjugate points is trivial, as can be understood from the following argument. Consider a trajectory in the neighborhood of the unstable periodic orbit and let it start with $\delta p_2' \neq 0$, $\delta q_2' = 0$. The successive crossings of the plane $\delta q_1 = 0$, i.e., the points $(\delta p_2'', \delta q_2'')$ after one, two, etc., periods, lie on a hyperbola which is centered at $\delta p_2 = 0$, $\delta q_2 = 0$. Therefore, one can never have $\delta q_2'' = 0$ as would be required if a conjugate point were to coincide with the initial point. Unstable periodic orbits are quite different from stable ones in this respect.

Formula (41), however, suffers from the same shortcoming which forced us to modify (36). If a conjugate point happens to be close to the "end point" of the periodic orbit, i.e., if it occurs shortly before or after one period, the starting formula (1) is poor. This fact manifests itself in a large amplitude because u is small, exactly as the amplitudes in (36) are large when nv is a multiple of 2π. As in the preceding section we argue now that the sinh $(nu/2)$ in the denominators of (41) should be replaced by a factor $\exp(-nu/2)$ in each term. This eliminates the singularity for small nu and preserves the exponential decay for large nu.

The contribution of a particular unstable periodic orbit to $\tilde{g}(E)$ is thereby modified to

$$-\frac{iT}{2\hbar}\sum_{n=1}^{\infty}\exp\left(in\frac{S}{\hbar}-inl\frac{\pi}{2}-n\frac{u}{2}\right). \tag{42}$$

For $u > 0$, this is a convergent geometric series which can be transformed exactly as (38) into

$$\frac{1}{2}\frac{\partial S}{\partial E}\left[\frac{1}{S-(\frac{1}{2}l\pi-\frac{1}{2}iu)\hbar}+\sum_{-\infty}^{+\infty}\left(\frac{1}{S-(2m\pi+\frac{1}{2}l\pi-\frac{1}{2}iu)\hbar}+\frac{1}{2m\pi\hbar}\right)+i\right]. \tag{43}$$

Again the factor $\frac{1}{2}$ disappears when we take into account that each periodic orbit can be traversed in two directions.

The response function $\tilde{g}(E)$ has simple poles of residue 1, but not on the real energy axis. If we calculate the density of states along the real energy axis from

$$\frac{i}{2\pi}[\tilde{g}(E+i0)-\tilde{g}(E-i0)] = \frac{\partial S}{\partial E}\sum_{m=-\infty}^{+\infty}\frac{\frac{1}{2}u\hbar}{[S-(m+\frac{1}{4}l)2\pi\hbar]^2+(\frac{1}{2}u\hbar)^2} \tag{44}$$

[cf. formula (10) of III], we find a series of broadened peaks. Their maxima are located at

$$S = (m + l/4)2\pi\hbar, \tag{45}$$

and their half-width is given by $u\hbar$, i.e., the distance of the two points along the S axis where the Lorentzian has dropped to half its peak value. The sharpness of the peaks, i.e., their width divided by their distance, is given by $u/2\pi$.

If our qualitative arguments are correct, we find the following situation: Stable periodic orbits give rise to δ-function singularities in the quasiclassical density of states at the energies which are given by the condition (39), whereas unstable periodic orbits give broadened peaks whose maxima are given by the condition (45). It was shown in III that there is, in general, a continuous background besides the δ-function singularities in the quasiclassical density of states. But the examples of III showed that all the eigenstates of Schrödinger's equation correspond only to the δ-function spikes, not the continuous background. It now appears, however, that the background cannot be neglected if the mechanical system has unstable periodic orbits.

This conclusion is, of course, rather unsatisfactory because it does not allow us to formulate a simple rule by which one could pick the approximate eigenvalues of the energy out of the quasiclassical density of states. On the other hand, even the broadened peaks associated with unstable periodic orbits are well defined as long as the exponent u is small compared to 2π. It may even be that the idea of a quasiclassical density of states was only useful in the derivation of the quantum conditions (39) and (45). The real question to be answered is the following: Is there a one-to-one correspondence between the eigenvalues of Schrödinger's equation and the energies given by the conditions (39) and (45)?

7. AN EXAMPLE FROM SEMICONDUCTOR PHYSICS

Let us consider an electron which is bound by an impurity of the donor type in one of the classical semiconductors, Si or Ge. Its wavefunction extends over many unit cells of the crystal lattice because the Coulomb attraction of the positively charged impurity is reduced by the large dielectric constant κ. Moreover, the effective kinetic energy is quadratic in the momenta, but since the conduction band minimum is not at a point of high crystal symmetry, this quadratic form is not a multiple of $|p|^2$. In an appropriate coordinate system the Hamiltonian is given by

$$H(p,q) = \frac{p_1^2}{2m_1}+\frac{p_2^2+p_3^2}{2m_2}-\frac{e^2}{\kappa|q|}, \tag{46}$$

where e is the electronic charge and the electronic mass tensor is given by m_1 and m_2. There is cylindrical symmetry around the 1 axis contrary to our previous hypothesis. But, in view of $m_1 > m_2$, the periodic orbits of lowest energy lie in a plane through the 1 axis, and we shall confine our attention to these. An additional complication has been neglected because we will consider only a single conduction band minimum, whereas there are six in Si and four in Ge. Each energy level is split according to this degree of freedom into sublevels. Any comparison of our results with experiments requires that we average over the corresponding sublevels.

The periodic orbit in the (q_1, q_2) plane satisfies the equations of motion

$$m_1\ddot{q}_1 = -e^2q_1/\kappa|q|^3, \quad m_2\ddot{q}_2 = -e^2q_2/\kappa|q|^3. \tag{47}$$

In order to obtain dimensionless variables, we use the following units of energy, length, and frequency:

$$E_0 = \frac{m_0e^4}{2\kappa^2\hbar^2}, \quad a_0 = \frac{\kappa\hbar^2}{m_0e^2}, \quad \omega_0 = \frac{m_0e^4}{\kappa^2\hbar^3}, \tag{48}$$

where $m_0 = (m_1 \cdot m_2)^{\frac{1}{2}}$. The new time parameter is $\tau = \omega_0 t$, and the new coordinates are

$$\xi(\tau) = q_1(t)/(\mu)^{\frac{1}{2}}a_0, \quad \eta(\tau) = q_2(t)/(\nu)^{\frac{1}{2}}a_0, \quad (49)$$

where $\mu = (m_2/m_1)^{\frac{1}{2}}$ and $\nu = (m_1/m_2)^{\frac{1}{2}}$. The normalized distance from the origin is called $\rho = (\mu\xi^2 + \nu\eta^2)^{\frac{1}{2}}$. The equations of motion become

$$\xi'' = \frac{\partial}{\partial\xi}\left(\frac{1}{\rho}\right), \quad \eta'' = \frac{\partial}{\partial\eta}\left(\frac{1}{\rho}\right), \quad (50)$$

where a prime indicates the derivative with respect to τ. The conservation of energy is expressed by the statement

$$\xi'^2 + \eta'^2 - 2/\rho = \epsilon, \quad (51)$$

where the normalized energy ϵ is given by E/E_0. The action integral becomes obviously

$$S = \int p\, dq = \hbar\int(\xi'^2 + \eta'^2)\, d\tau. \quad (52)$$

The search of a simple periodic orbit follows exactly the classic work of Hill in the restricted 3-body problem. The details can be found in the textbook by Brouwer and Clemence.[7] The orbit is expanded in a Fourier series of $\omega\tau$, where the normalized frequency ω remains to be determined. With $z = \exp(i\omega\tau)$, the symmetry of the orbit with respect to the ξ and η axes leads to the expressions

$$\xi + i\eta = \alpha w, \quad w = \sum_{-\infty}^{+\infty}\alpha_j z^{2j+1},$$

$$\xi - i\eta = \alpha\bar{w}, \quad \bar{w} = \sum_{-\infty}^{+\infty}\beta_j z^{2j+1}, \quad (53)$$

where $\beta_j = \alpha_{-j-1}$ and these coefficients are real. The scale factor α in front of w and $\bar{w}$ is necessary because the coefficients α_j and β_j will be obtained from a fixed algorithm to be described later.

The equations of motion can be written in terms of the operator $D = z \cdot d/dz$. Furthermore, in view of $\xi = \frac{1}{2}\alpha(w + \bar{w})$ and $\eta = \alpha(w - \bar{w})/2i$, we introduce

$$\Omega = \alpha^{-3}\omega^{-2}[\tfrac{1}{4}\mu(w + \bar{w})^2 - \tfrac{1}{4}\nu(w - \bar{w})^2]^{-\frac{1}{2}}, \quad (54)$$

which gives the equations of motion

$$D^2w + 2\frac{\partial\Omega}{\partial\bar{w}} = 0, \quad D^2\bar{w} + 2\frac{\partial\Omega}{\partial w} = 0, \quad (55)$$

with the conservation of energy

$$Dw \cdot D\bar{w} + 2\Omega + \epsilon/\alpha^2\omega^2 = 0. \quad (56)$$

These equations for an electron with anisotropic mass tensor in a Coulomb field are entirely analogous to Hill's equation for the restricted 3-body problem.

Since Ω is homogeneous in w and $\bar{w}$ of degree -1, we find immediately after multiplying the first Eq. (55) with $\bar{w}$ and the second Eq. (55) with w, and after inserting the value of Ω from (56), that

$$\bar{w} \cdot D^2w + w \cdot D^2\bar{w} + Dw \cdot D\bar{w} + \epsilon/\alpha^2\omega^2 = 0. \quad (57)$$

Again, we can compare $w \cdot D^2\bar{w} - \bar{w} \cdot D^2w$ with $w \cdot D^2w - \bar{w} \cdot D^2\bar{w}$ as obtained from (55) and (54). They differ only by a factor $(\mu - \nu)/(\nu + \mu)$. Therefore, we get the equation

$$w \cdot D^2\bar{w} - \bar{w} \cdot D^2w + m(w \cdot D^2w - \bar{w} \cdot D^2\bar{w}) = 0, \quad (58)$$

where we used $m = (\nu - \mu)/(\nu + m) = (m_1 - m_2)/(m_1 + m_2)$.

The ratio m is small when m_1 and m_2 are nearly equal. Although this is by no means the case for either Si or Ge, one can easily construct an algorithm for obtaining w and $\bar{w}$ from (57) and (58) which corresponds to a power series in m and which converges quite well for values of m close to its maximum 1. This algorithm is described in a paper by Eckert and Eckert.[8] The details are given in Appendix A. We obtain in this manner the coefficients α_j in (53) for $\epsilon = -\alpha^2\omega^2$ and any arbitrary m by numerical iteration, such that (57) and (58) are satisfied. Moreover, we determine a number $\gamma = \alpha^3\omega^2$ such that Eqs. (55) hold in addition to (57) and (58). The value of γ depends only on m and is always near 1. The relation $\gamma = \alpha^3\omega^2$ is the analog of Kepler's third law.

The value of the action integral S around the closed orbit is given according to (52), in units of $\hbar$, by

$$\oint(\xi'^2 + \eta'^2)\, d\tau = (-\epsilon) \cdot \oint d\tau = -\epsilon\frac{2\pi}{\omega}. \quad (59)$$

The virial theorem has been used. It can, indeed, be derived from (50) and (51) that the average kinetic energy over a periodic orbit equals $|\epsilon|$. The relation between ω and ϵ follows from the two equations $\epsilon = -\alpha^2\omega^2$ and $\gamma = \omega^2\alpha^3$. Thus we find from (59) and (52) that

$$\oint p\, dq = 2\pi\hbar \cdot \gamma/(|\epsilon|)^{\frac{1}{2}}. \quad (60)$$

The values of γ as a function of m are graphed in Fig. 1. It is remarkable how close to 1 they are. If γ is approximated by 1, the action integral (60) assumes the classical expression of Bohr for the hydrogen atom provided that we use the mass $m_0 = (m_1m_2)^{\frac{1}{2}}$ and the charge reduced by $(\kappa)^{\frac{1}{2}}$.

In order to apply the quantization condition of the two preceding sections, we have to examine the

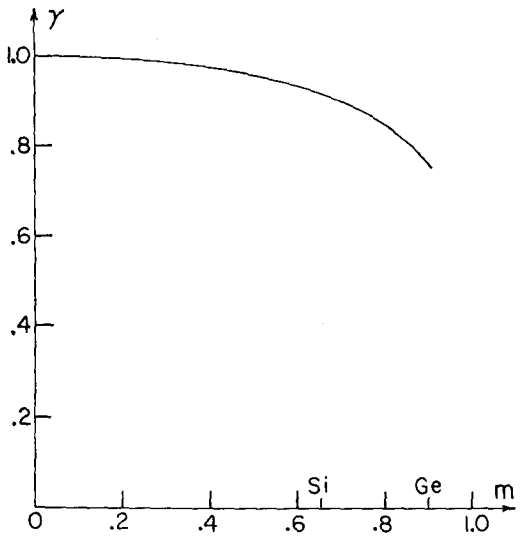

FIG. 1. The constant γ, which appears in "Kepler's third law," $\gamma = \omega^2\alpha^3$, is plotted as a function of the anisotropy parameter $m = (m_1 - m_2)/(m_1 + m_2)$. The approximate eigenvalues of the energy are given by "Balmer's formula," $\epsilon = -\gamma^2/n^2$, in the normalization (48).

stability of the periodic orbit in the (ξ, η) plane. This investigation again follows Hill's classical work on the motion of the perigee of the moon. The details are found in the textbook of Brouwer and Clemence.[7] If the periodic orbit is given by the Fourier series $\xi_0(\tau)$ and $\eta_0(\tau)$, we are now looking for a more general solution of (50) in the form

$$\xi = \xi_0 + \delta\xi, \quad \eta = \eta_0 + \delta\eta. \tag{61}$$

The expressions are inserted into (50) and (51), but only linear terms in the deviations $\delta\xi$ and $\delta\eta$ will be kept. If we write

$$\delta\xi = \lambda\frac{d\xi_0}{d\tau} - \sigma\frac{d\eta_0}{d\tau}, \quad \delta\eta = \sigma\frac{d\xi_0}{d\tau} + \lambda\frac{d\eta_0}{d\tau}, \tag{62}$$

we immediately get a particular solution $\lambda = 1, \sigma = 0$, which is of no interest. The first task is, therefore, to find one second-order linear differential equation in σ alone, rather than two coupled second-order equations in both λ and σ.

This process of elimination is performed in detail by Brouwer and Clemence[7] who use the Fourier series (53) in the form

$$\delta w = (i\lambda - \sigma)Dw_0, \quad \delta\bar{w} = (i\lambda + \sigma)D\bar{w}_0, \tag{63}$$

together with the Eqs. (55) and (56). One ends up with Hill's equation

$$D^2\psi - \theta\psi = 0, \tag{64}$$

where ψ is related to σ through

$$\psi = \sigma \cdot (-Dw_0 \cdot D\bar{w}_0)^{\frac{1}{2}}, \tag{65}$$

and the function θ can be expressed in many equivalent forms. The most suitable for our purpose is given by

$$\theta = \frac{3}{(Dw_0D\bar{w}_0)^2}\left(-\frac{\partial\Omega}{\partial w}Dw_0 + \frac{\partial\Omega}{\partial\bar{w}}D\bar{w}_0\right)^2 + \frac{1}{Dw_0 \cdot D\bar{w}_0}\left(\frac{\partial^2\Omega}{\partial w^2}(Dw_0)^2 - 2\frac{\partial^2\Omega}{\partial w\partial\bar{w}}Dw_0D\bar{w}_0 + \frac{\partial^2\Omega}{\partial\bar{w}^2}(D\bar{w}_0)^2\right), \tag{66}$$

where the derivatives of Ω are to be evaluated for $w = w_0$ and $\bar{w} = \bar{w}_0$. If we insert the expression (54) for Ω with $4\rho_0^2 = \mu(w_0 + \bar{w}_0)^2 - \nu(w_0 - \bar{w}_0)^2$, we find that

$$\begin{aligned}\theta = {} & (4\gamma\rho_0^3Dw_0D\bar{w}_0)^{-1}\{2(\mu + \nu)Dw_0D\bar{w}_0 \\ & + (\nu - \mu)[(Dw_0)^2 + (D\bar{w}_0)^2]\} \\ & + (3/16\gamma\rho_0^5Dw_0D\bar{w}_0)[1 + (\gamma\rho_0Dw_0D\bar{w}_0)^{-1}] \\ & \times [(\mu + \nu)(\bar{w}_0Dw_0 - w_0D\bar{w}_0) \\ & + (\mu - \nu)(w_0Dw_0 - \bar{w}_0D\bar{w}_0)]^2.\end{aligned} \tag{67}$$

The evaluation of θ as a power series in ζ is straightforward. With $\epsilon = -\alpha^2\omega^2$ and $\gamma = \alpha^3\omega^2$, it follows from (56) that $\frac{1}{2}(1 - Dw_0D\bar{w}_0) = 1/\gamma\rho_0$, which gives the Fourier series for ρ_0^{-1}. The Fourier expansion of $1/Dw_0D\bar{w}_0$ can be obtained by a simple iteration scheme because the largest term in $Dw_0D\bar{w}_0$ is the constant -1. The details are described in Appendix B.

Hill's equation (64) can, of course, be solved by the classical procedure described in the standard texts. The Fourier expansion for θ does not converge very fast, however, and it is simpler to integrate (64) numerically to obtain the stability index in this way. Since θ is even in τ, we can choose the functions $\psi_1 = f(\tau)\exp(\beta\tau)$ and $\psi_2 = f(-\tau)\exp(-\beta\tau)$ as the independent solutions of (64), where $f(\tau)$ has the periodicity of θ and is normalized to $f(0) = 1$. The solution $\psi = \frac{1}{2}(\psi_1 + \psi_2)$ is characterized by $\psi(0) = 1$ and $\psi'(0) = 0$. Therefore, if one integrates Eq. (64) numerically with these initial conditions, he finds that $\psi(2\pi/\omega) = \cosh(2\pi\beta/\omega)$.

From this relation one obtains immediately the value of $u = 2\pi\beta/\omega$, which determines the stability of the orbit and the width of the peak in the density of states. The values of u are plotted in Fig. 2.

The numerical solution of (64) shows that the periodic orbit we have obtained is unstable for all values of m. Moreover, the function ψ goes through 0 twice when τ increases from 0 to $2\pi/\omega$, yielding, therefore, two conjugate points along the orbit. Thus we find, with the help of (60) for the 2-dimensional

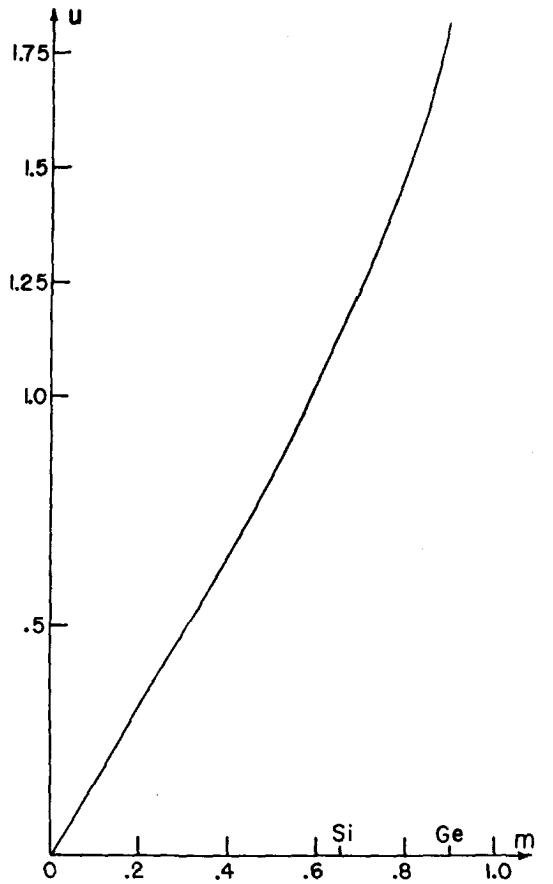

FIG. 2. The instability exponent u is plotted as a function of the anisotropy parameter $m = (m_1 - m_2)/(m_1 + m_2)$. The width of the peaks in the classical density of states is $u\hbar$ whereas their separation is $2\pi\hbar$ on the action scale, rather than the usual energy scale.

case of the anisotropic Kepler problem, the quantum condition

$$\gamma/(|\epsilon|)^{\frac{1}{2}} = n + \tfrac{1}{2} \tag{68}$$

and the approximate eigenvalues of the energy

$$\epsilon = -\gamma^2/(n + \tfrac{1}{2})^2. \tag{69}$$

In order to compare our method of approximation with the exact results, we have to examine the 3-dimensional case which results from the Hamiltonian (46). The two Eqs. (47) have to be complemented by a third equation which is given by

$$m_2\ddot{q}_3 = -e^2 q_3/\kappa\,|q|^3. \tag{70}$$

The coordinate q_3 is now reduced exactly as q_2 by

$$\zeta(\tau) = q_3(t)/(\nu)^{\frac{1}{2}}a_0, \tag{71}$$

which leads to the reduced equation of motion

$$\zeta'' = \partial(1/\rho)/\partial\zeta, \tag{72}$$

where $\rho = (\mu\xi^2 + \nu\eta^2 + \nu\zeta^2)^{\frac{1}{2}}$. The energy equation (51) contains an additional term ζ'^2, and so does the expression (52) for the action integral.

If ζ is considered as a small deviation from the periodic orbit in the (ξ, η) plane, it satisfies the linear second-order equation

$$\frac{d^2\zeta}{d\tau^2} + (\mu\xi^2 + \nu\eta^2)^{-\frac{3}{2}} \cdot \zeta = 0, \tag{73}$$

where $\xi(\tau)$ and $\eta(\tau)$ are known functions of period $2\pi/\omega$. The solution $\zeta_1 = \eta(\tau)$ to (73) is already known so that one can easily write down a second solution of (73), namely

$$\zeta_2(\tau) = \eta(\tau) \cdot \int_{\pi/2\omega}^{\tau} \frac{d\tau'}{\eta^2(\tau')}. \tag{74}$$

This solution is characterized by $\zeta_2(0) \neq 0$ and $\zeta_2'(0) = 0$, whereas $\zeta_1(0) = 0$ and $\zeta_1'(0) \neq 0$. Both ζ_1 and ζ_2 have the periodicity $2\pi/\omega$.

The third dimension in the anisotropic Kepler problem introduces two more conjugate points along the periodic orbit which has been studied so far. The orbit is neither stable nor unstable with respect to this degree of freedom, a fact which follows immediately from the cylindrical symmetry around the ξ axis. Thus, one arrives at the approximate eigenvalues of the energy

$$\epsilon = -\gamma^2/(n + 1)^2, \tag{75}$$

with $n = 0, 1, 2, \cdots$ for the particular periodic orbit in a plane through the ξ axis. The similarity with the Balmer formula is obvious, but one has to remember that the value of γ depends on the periodic orbit as well as on the anisotropy m. Another periodic orbit will generally have a different value of γ in (60), and the denominator $(n + 1)^2$ in (75) may have to be replaced by $(n + \delta)^2$, where δ is determined by the stability and the number of conjugate points.

8. COMPARISON WITH "EXPERIMENT"

Excited states of an electron in a donor impurity are known from infrared spectroscopy in a number of cases, particularly for Li, P, As, Sb, and Bi in Si and Ge.[9] The effective masses at the conduction band minima are known from cyclotron resonance to 0.1%, and the dielectric constant from the long wavelength index of refraction. Experiments seem to have outstripped our theoretical understanding.

The present investigation neglects both the existence of more than one minimum in the conduction band and the atomic structure of the donor impurity. It would, therefore, be foolish to attempt a direct comparison of the experimental results with the approximate eigenvalues of the energy (75). Since we are studying the use of classical mechanics in the

solution of Schrödinger's equation, we shall compare our results with the recent work of Faulkner,[4] who has obtained approximate eigenstates for the quantum mechanical Hamiltonian (46) by using finite linear combinations of hydrogen wavefunctions. In this manner he gets approximate eigenvalues of the energy from the ground state up to the sixth atomic shell.

The Hamiltonian (46) has cylindrical symmetry around the 1 axis so that its eigenstates can be characterized by their angular momentum around the 1 axis. Since our periodic orbit lies in a plane through the 1 axis, our approximate energies (75) correspond to eigenstates with zero angular momentum around the 1 axis. The Hamiltonian (46) has also reflection symmetry with respect to the (2, 3) plane so that its eigenstates can be further characterized by their reflection properties. There are even and odd states with respect to the reflection on the (2, 3) plane, but our approximation does not distinguish between them.

Thus, we are left with the set of states which Faulkner calls $1S$, $2S$, $2P_0$, $3S$, $3P_0$, $3D_0$, $4S$, $4P_0$, $4D_0$, $4F_0$, etc., to make a comparison. These designations are unambiguous for small anisotropy, i.e., for small values of m, because the anisotropy causes only small admixtures to the corresponding pure hydrogen states. Actually, a first-order perturbation theory gives the splitting of the hydrogen levels due to the anisotropy, as will be indicated presently. If we use the normalization (48), i.e., the Hamiltonian which results from (51), the perturbation is simply the deviation of $1/\rho$ from $(\xi^2 + \eta^2 + \zeta^2)^{-\frac{1}{2}}$. In terms of $m = (m_1 - m_2)/(m_1 + m_2)$, we obviously have $\mu = 1 - m + \cdots$ and $\nu = 1 + m + \cdots$. Thus, we find that

$$\rho^{-1} = (\xi^2 + \eta^2 + \zeta^2)^{-\frac{1}{2}} \times [1 + \tfrac{1}{2}m(\cos^2\vartheta - \sin^2\vartheta) + \cdots], \quad (76)$$

where ϑ is the angle with the ξ axis. The second term on the right-hand side is the perturbation. Its matrix elements are small for two hydrogen wavefunctions of same principal quantum number n, but different azimuthal quantum number l. Thus, there is no need for degenerate perturbation theory, and one finds the eigenvalues of the energy to first order,

$$\epsilon = -\frac{1}{n^2}\left(1 + \frac{m}{(2l-1)(2l+3)} + \cdots\right), \quad (77)$$

in the normalization (48).

This result has to be contrasted with a calculation of our periodic orbit to lowest order in m. Equations (57) and (58) give immediately the approximation

$$w = z(1 + \tfrac{1}{8}mz^2 - \tfrac{7}{24}mz^2 + \cdots) \quad (78)$$

for $\epsilon = -\alpha^2\omega^2$. The number γ equals 1 with corrections of order m^2. For completeness sake, we list

$$\theta = 1 - \tfrac{1}{2}m(z^2 + z^{-2}) + \cdots \quad (79)$$

to lowest order in m, which is to be inserted into Hill's equation (64). It follows that the stability exponent is given by

$$u = \tfrac{1}{2}m\pi + \cdots, \quad (80)$$

with the help of Hill's determinant.

To lowest order in m, we find, therefore, the approximate energies $\epsilon = -1/n^2$ with $n = 1, 2, \cdots$. This last formula coincides with (77) in the limit of very large l. It is now conjectured that the energies (75) always correspond to an eigenstate with the largest possible expectation value of the azimuthal angular momentum. In Faulkner's nomenclature, the series of energies (75) corresponds to the levels $1S$, $2P_0$, $3D_0$, $4F_0$, $5G_0$, $6H_0$, where the lower index 0 indicates the vanishing magnetic quantum number. The periodic orbits which were obtained in the preceding section are simply closed, symmetric curves in a plane through the ξ axis and never get close to the origin. A periodic orbit of low average angular momentum, however, would either be asymmetric or come close to the origin.

In order to compare the numerical results of Faulkner with formula (75), it is necessary to give them a somewhat different interpretation. It is clear from (77) that for a given n the S level is always highest whereas the P level is lowest, and the other levels have increasing energies with increasing l. This level scheme is intuitively obvious from the expression (51) for the energy because the potential energy $-2/\rho$ is higher at some point on the η axis than it is at a point on the ξ axis which is at the same distance from the origin. An S state tends to have the same amplitude along the ξ and η axes, while a P_0 state has a vanishing amplitude along the η axis because it is an odd function of ξ. The states with larger average angular momentum do not avoid the η axis as well as the P_0 state, but they surely have most of their amplitude along the ξ axis.

Unfortunately, this simple argument was not recognized by Faulkner, and he was, therefore, led to mislabel his results, although his numbers seem correct. A more subtle and, in a certain sense, arbitrary consideration has to do with the crossing of levels. The formula (77) suggests that the P_0 level for some value of n might cross the S level for a lower value of n as m increases. Since P_0 is odd and S is even with respect to reflection on the (η, ζ) plane, there will be a real crossing. If m increases further, however, it might

happen that even the D_0 level tries to drop below the S level which belongs to a smaller value of n. The two curves will not cross on a plot of their energy versus m. Rather, the appearance of the wavefunction will change continuously from a D_0 type into an S type and vice versa. If this change occurs fairly quickly on the m scale, it would appear entirely reasonable to change the label accordingly, although the point at which the labels are switched is necessarily somewhat arbitrary.

The results of Faulkner's calculations for Si and Ge are presented in Table I together with the energies resulting from (75). Our interpretation of the levels is given in the left-hand column whereas Faulkner's labels are given to the right of his numerical results. The masses are $m_1 = 0.916\ m_e$ and $m_2 = 0.1905\ m_e$ with $m = 0.656$ and $\kappa = 11.4$ for Si, and $m_1 = 1.588\ m_e$ and $m_2 = 0.0815\ m_e$ with $m = 0.899$ and $\kappa = 15.36$ for Ge, where m_e is the mass of the free electron. Notice that our approximate values for Si are always above the energy of the level with the largest angular momentum, but below the S level. In the case of Ge this rule seems to hold only for the higher levels $n = 4$, 5, and 6, but not for $n = 2$ and 3. Of course, Ge is further away from the simple hydrogen model than Si so that both our interpretation of levels in Ge and Faulkner's calculations become more questionable.

TABLE I. The results of Faulkner's calculations based on expanding the electron wavefunctions around the donor impurity into a sum of hydrogenlike terms, are compared with the quasiclassical approximation. Column 1 gives our interpretation of the various levels. Columns 2 and 5 give our approximate results for certain special levels, while columns 3 and 6 give Faulkner's results for all levels. Columns 4 and 7 give Faulkner's designations of these levels for silicon and germanium.

Level	Si			Ge		
$1s$	36.81	31.27	$1S$	11.92	9.81	$1S$
$2p_0$	9.20	11.51	$2P_0$	2.98	4.74	$2P_0$
$2s$		8.83	$2S$		3.52	$2S$
$3p_0$		5.48	$3P_0$		2.56	$3P_0$
$3d_0$	4.09	4.75	$3S$	1.32	2.01	$3S$
$3s$		3.75	$3D_0$		1.33	$3D_0$
$4p_0$		3.33	$4P_0$		1.67	$4P_0$
$4d_0$		2.85	$4S$		1.17	$4S$
$4f_0$	2.30	2.33	$4F_0$	0.75	0.80	$5F_0$
$4s$		2.11	$4D_0$		0.72	$5S$
$5p_0$		2.23	$5P_0$		1.16	$4F_0$
$5d_0$		1.87	$5S$		0.87	$4D_0$
$5f_0$		1.62	$5F$		0.55	$6F_0$
$5g_0$	1.47	1.52	$5D_0$	0.48	0.61	$5D_0$
$5s$		1.38	$5G$		0.53	$5G_0$
$6p_0$		1.52	$6P_0$		0.84	$5P_0$
$6d_0$						
$6f_0$		1.20	$6F_0$		0.61	$6P_0$
$6g_0$						
$6h_0$	1.02	1.10	$6H_0$	0.33	0.40	$6H_0$
$6s$						

The anisotropic Kepler problem was discussed only to illustrate the use of periodic orbits in approximating the eigenvalues of a nonseparable Hamiltonian. Furthermore, we have considered only a type of periodic orbits which is sufficiently close to a circle and allows the use of methods from lunar theory for its computation. If we want to find the classical analog of the levels in Table I, where the columns two and five are blank, we have to search for other types of periodic orbits in the (ξ, η) plane. These will have more complicated shapes; in particular, these orbits will intersect themselves and may lack any symmetry with respect to the ξ axis or the η axis. Hill's method for finding periodic orbits can probably not be applied to them, and one has to rely on the more general methods of Poincaré and his successors. Such an investigation will concentrate mainly on finding certain classes of periodic orbits and characterizing them in terms of their topology and stability. But it should be kept in mind that these periodic orbits are eventually to be interpreted in terms of energy levels of a corresponding Schrödinger equation. The above example suggests that such an interpretation yields not only a simple approximation to the eigenvalues of the energy, but provides a better understanding of the solutions of Schrödinger's equation.

ACKNOWLEDGMENTS

The author wishes to express his gratitude to Professor W. J. Eckert, who has spent a lot of time trying to convey the essentials of celestial mechanics and who suggested the Fourier expansion method for finding the periodic orbits. Dr. A. Onton was extremely helpful in giving the author the relevant references, both theoretical and experimental, about electron spectra in donor impurities. Finally, the author wants to thank his colleagues at the Watson Laboratory for numerous informal discussions and hints, particularly J. Fortoul, N. D. Lang, and P. J. Price.

APPENDIX A

For $\epsilon = -\alpha^2\omega^2$ and $m = 0$, Eqs. (57) and (58) have the trivial solution $w = z$, i.e., $\alpha_0 = 1$ and $\alpha_j = 0$ for $j \neq 0$. This solution will be taken as the zero-order approximation for w when $m \neq 0$. The solution of (57) and (58) will be generated by an iteration procedure which uses only the result of the preceding step. Thus, let us assume that the nth iteration has given the power series w. The $(n + 1)$th iteration will then provide a correction δw and the new approximation $w + \delta w$.

Since δw is small (in some sense) compared to w, we neglect all nonlinear terms in δw when we insert $w + \delta w$ into (57) and (58). This would leave terms of the type $\bar{w} \cdot \delta w$, $w \cdot \delta \bar{w}$, etc. However, we shall go one step further in the effort to simplify the determination of δw. In all terms of the type $\bar{w}\delta w$, $w\delta\bar{w}$, etc., we shall replace w by z, its zero-order approximation, and $\bar{w}$ by z^{-1}. Moreover, we shall neglect all terms where δw or $\delta\bar{w}$ gets multiplied by m.

If we use the power series U and V defined by

$$\delta w = zU, \quad \delta\bar{w} = z^{-1}V, \tag{A1}$$

the equations to be solved at each step of the iteration become

$$(D^2 + D + 1)U + (D^2 - D + 1)V = -\bar{w} \cdot D^2 w - w \cdot D^2\bar{w} - Dw \cdot D\bar{w} - \epsilon/\alpha^2\omega^2, \tag{A2}$$

$$-(D^2 + 2D)U + (D^2 - 2D)V = -w \cdot D^2\bar{w} + \bar{w} \cdot D^2 w - m(w \cdot D^2 w - \bar{w} \cdot D\bar{w}). \tag{A3}$$

The right-hand sides are just the "residues" from Eqs. (57) and (58), i.e., the quantities by which w and $\bar{w}$ fail to satisfy the Eqs. (57) and (58).

The functions U and V are power series in z whose general form can be obtained from (A1) and (53). Thus, we can write

$$U = \sum_{-\infty}^{+\infty} u_j z^{2j}, \quad V = \sum_{-\infty}^{+\infty} v_j z^{2j}, \tag{A4}$$

with $v_j = u_{-j}$. Equations (A2) and (A3) become

$$(4j^2 + 2j + 1)u_j + (4j^2 - 2j + 1)v_j = r_j, \tag{A5}$$

$$-(4j^2 + 4j)u_j + (4j^2 - 4j)v_j = s_j, \tag{A6}$$

if the coefficients of equal powers in z are matched and if the right-hand sides of (A2) and (A3) are written as $\sum r_j z^{2j}$ and $\sum s_j z^{2j}$. It can easily be checked that $r_j = r_{-j}$ and $s_j = -s_{-j}$, so that the Eqs. (A5) and (A6) are compatible with the relation $v_j = u_{-j}$. The determinant of the coefficients of u_j and v_j is given by $8j^2(4j^2 - 1)$.

The evaluation of the residues involves no more than summation, multiplication, and differentiation of power series. All these operations are easily programmed on a computer. Beyond a certain value of m, the convergence of the iteration scheme is not monotonic, but oscillatory and very slow. It is enough to take $w + \delta w/2$ as the new approximation at every fifth iteration, say, to prevent this oscillatory approach to the final result.

The solution of (57) and (58) implies the existence of a potential like Ω as given by (54) up to an arbitrary factor. This factor is fixed by requiring either (55) or (56) to be satisfied at one particular value of τ, e.g., at $\tau = 0$ which implies $z = 1$. If we abbreviate $\gamma = \omega^2\alpha^3$, we find from (56) that

$$\frac{1}{\gamma} = \tfrac{1}{4}(\mu)^{\frac{1}{2}}(1 - DwD\bar{w})(w + \bar{w})\big|_{z=1}, \tag{A7}$$

where we have put $\epsilon = -\alpha^2\omega^2$ all along.

APPENDIX B

After w has been obtained with the help of the iteration procedure of Appendix A, the periodic function θ has to be evaluated from the formula (67). Since $1/\rho$ follows from the conservation of energy, θ could be calculated using only the summation, multiplication, and differentiation of power series if it were not for the division by $Dw_0D\bar{w}_0$. This last quantity is essentially the kinetic energy, and remains, therefore, positive for all values of z on the unit circle.

$Dw_0D\bar{w}_0$ contains only the even powers of z, and its coefficients are symmetric; i.e., if we write

$$W = -Dw_0 \cdot D\bar{w}_0 = \sum_{-\infty}^{+\infty} W_j z^{2j}, \tag{B1}$$

we find that $W_j = W_{-j}$ and these coefficients are real. Moreover, $W_0 = 1$ because the average kinetic energy equals $-\epsilon/\alpha^2\omega^2$ according to (56) and the virial theorem. For small values of m the coefficients W_j are small for $j \neq 0$. Therefore, one expects that the power series expansion of $F = 1/W$ has small coefficients except for the term $j = 0$.

The equation $F \cdot W = 1$ is then solved by iteration exactly as the Eqs. (57) and (58) were solved in Appendix A. The zero-order approximation for F is 1. Again we assume that the nth iteration has yielded F, and we use the $(n + 1)$th iteration to obtain a correction δF. But, in evaluating $(F + \delta F)W = 1$, we replace $\delta F \cdot W$ simply by δF as if W were equal to 1. Thus we find the correction $\delta F = 1 - FW$ and the input $F + \delta F$ for the next iteration.

This procedure works only if W is close to 1. If such is not the case, however, the power series W has to be inverted in steps. Thus, let us call

$$W^{(l)} = 1 - (l/L) + (l/L)W, \tag{B2}$$

where L is some positive integer and $1 \leq l \leq L$. If L is large enough, $W^{(1)}$ satisfies the assumption of the preceding paragraph, and we can find $F^{(1)}$ to satisfy the equation $F^{(1)} \cdot W^{(1)} = 1$. Now, suppose that we have found $F^{(l)}$ to satisfy the equation $F^{(l)}W^{(l)} = 1$ and that we have an nth approximation $F^{(l)} + \delta F$

toward $F^{(l+1)}$ which is required to satisfy $F^{(l+1)} \cdot W^{(l+1)} = 1$. The $(n+1)$th approximation to $F^{(l+1)}$ would be given by $F^{(l)} + F + \delta F$, where we try to satisfy

$$(F^{(l)} + F + \delta F)W^{(l+1)}$$
$$= (F^{(l)} + F)W^{(l+1)} + \delta F \cdot W^{(l)} + \delta F(W^{(l)} - 1) = 1.$$

But, we now assume $\delta F \cdot W^{(l)}$ to be small, and we use our knowledge of $F^{(l)}$ to write

$$\delta F = F^{(l)}[1 - (F^{(l)} + F)W^{(l+1)}]. \qquad \text{(B3)}$$

In this manner $F^{(l+1)}$ is found by iteration, and ultimately so is $F^{(L)}$.

* Present address: IBM Watson Research Center, Yorktown Heights, New York 10598.

[1] M. C. Gutzwiller, J. Math. Phys. **8**, 1979 (1967); **10**, 1004 (1969); **11**, 1791 (1970). These papers will be referred to as I, II, and III.
[2] S. Smale, Bull. Am. Math. Soc. **73**, 747 (1967).
[3] W. Kohn and J. M. Luttinger, Phys. Rev. **96**, 1488 (1954).
[4] R. A. Faulkner, Phys. Rev. **184**, 713 (1969).
[5] S. Smale, in *Differential and Combinatorial Topology*, Symposium in Honor of Marston Morse, Stewart S. Cairns, Ed. (Princeton U.P., Princeton, N.J.), pp. 63–80.
[6] For a discussion of the monodromy matrix, cf. L. A. Pars, *A Treatise on Analytical Dynamics* (Heinemann, London, 1965), p. 461. The area preserving map near a periodic orbit is discussed in C. Siegel, *Vorlesungen über Himmelsmechauik* (Springer-Verlag, Berlin, 1956), p. 131.
[7] D. Brouwer and G. M. Clemence, *Methods of Celestial Mechanics* (Academic, New York, 1961), p. 340.
[8] W. J. Eckert and D. A. Eckert, Astron. J. **72**, 1299 (1967).
[9] For references about recent experimental work, cf. Ref. 4.

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 12, NUMBER 3 MARCH 1971

Scattering of Waves in Many Dimensions

I. ADAWI
Physics Department, University of Missouri at Rolla, Rolla, Missouri 65401

(Received 4 September 1970)

Scattering by a spherical potential is discussed in all dimensions by one formulation using the partial-wave expansion method. The optical theorem relating the total scattering cross section σ to the forward scattering amplitude $f(0)$ is derived.

We wish to derive a formula relating the total scattering cross section to the forward-scattering amplitude which holds in all dimensions. The procedure is to use Gegenbauer's expansion[1] of a plane wave in N dimensions in terms of partial waves appropriate for the space under consideration. This allows us to treat scattering in all dimensions on equal footing, without resort to special formulations, and reveals the similarities of wave motion in all dimensions.

The stationary-state Schrödinger equation is

$$[\Delta + k^2 - U(x)]\psi = 0, \qquad (1)$$

where Δ is the Laplacian operator in N dimensions, $k^2 = (2m/\hbar^2)E$, E is the energy, m is the mass of the particle, and $U(x) = (2m/\hbar^2)V(x)$, where $V(x)$ is the potential and x is the length of the N-dimensional position vector $\mathbf{x}$. It is thus assumed that the potential is spherically symmetric. We are interested in the scattering of a wave incident along the z axis, and ψ will depend on x and θ where $z = x\cos\theta$. The Laplacian takes the form

$$\Delta = \frac{\partial^2}{\partial x^2} + \frac{N-1}{x}\frac{\partial}{\partial x} + \frac{N-2}{x^2}\cot\theta\frac{\partial}{\partial\theta} + \frac{1}{x^2}\frac{\partial^2}{\partial\theta^2}. \qquad (2)$$

The wave equation (1) is separable, and the solutions are partial waves of the form $f_n(x)C_n^{\alpha}(\cos\theta)$, where $f_n(x)$ is the radial solution and $C_n^{\alpha}(\cos\theta)$ is the N-dimensional spherical harmonic of Gegenbauer, in which n is the degree of the polynomial and $\alpha = \frac{1}{2}N - 1$.

These harmonics are defined by the generating function[2,3]

$$(1 - 2h\cos\theta + h^2)^{-\alpha} = \sum_{n=0}^{\infty} h^n C_n^{\alpha}(\cos\theta) \qquad (3)$$

and satisfy the differential equation

$$\left(\frac{d^2}{d\theta^2} + (N-2)\cot\theta\frac{d}{d\theta} + n(N+n-2)\right)C_n^{\alpha}(\cos\theta) = 0. \qquad (4)$$

By substituting (4) and (2) into (1) we obtain the radial equation

$$\left(\frac{d^2}{dx^2} + \frac{2\alpha+1}{x}\frac{d}{dx} + k^2 - n(n+2\alpha)x^{-2} - U(x)\right)f_n(x) = 0. \qquad (5)$$

toward $F^{(l+1)}$ which is required to satisfy $F^{(l+1)} \cdot W^{(l+1)} = 1$. The $(n+1)$th approximation to $F^{(l+1)}$ would be given by $F^{(l)} + F + \delta F$, where we try to satisfy

$$(F^{(l)} + F + \delta F)W^{(l+1)}$$
$$= (F^{(l)} + F)W^{(l+1)} + \delta F \cdot W^{(l)} + \delta F(W^{(l)} - 1) = 1.$$

But, we now assume $\delta F \cdot W^{(l)}$ to be small, and we use our knowledge of $F^{(l)}$ to write

$$\delta F = F^{(l)}[1 - (F^{(l)} + F)W^{(l+1)}]. \qquad \text{(B3)}$$

In this manner $F^{(l+1)}$ is found by iteration, and ultimately so is $F^{(L)}$.

* Present address: IBM Watson Research Center, Yorktown Heights, New York 10598.

[1] M. C. Gutzwiller, J. Math. Phys. **8**, 1979 (1967); **10**, 1004 (1969); **11**, 1791 (1970). These papers will be referred to as I, II, and III.
[2] S. Smale, Bull. Am. Math. Soc. **73**, 747 (1967).
[3] W. Kohn and J. M. Luttinger, Phys. Rev. **96**, 1488 (1954).
[4] R. A. Faulkner, Phys. Rev. **184**, 713 (1969).
[5] S. Smale, in *Differential and Combinatorial Topology*, Symposium in Honor of Marston Morse, Stewart S. Cairns, Ed. (Princeton U.P., Princeton, N.J.), pp. 63–80.
[6] For a discussion of the monodromy matrix, cf. L. A. Pars, *A Treatise on Analytical Dynamics* (Heinemann, London, 1965), p. 461. The area preserving map near a periodic orbit is discussed in C. Siegel, *Vorlesungen über Himmelsmechauik* (Springer-Verlag, Berlin, 1956), p. 131.
[7] D. Brouwer and G. M. Clemence, *Methods of Celestial Mechanics* (Academic, New York, 1961), p. 340.
[8] W. J. Eckert and D. A. Eckert, Astron. J. **72**, 1299 (1967).
[9] For references about recent experimental work, cf. Ref. 4.

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 12, NUMBER 3 MARCH 1971

Scattering of Waves in Many Dimensions

I. ADAWI
Physics Department, University of Missouri at Rolla, Rolla, Missouri 65401

(Received 4 September 1970)

Scattering by a spherical potential is discussed in all dimensions by one formulation using the partial-wave expansion method. The optical theorem relating the total scattering cross section σ to the forward scattering amplitude $f(0)$ is derived.

We wish to derive a formula relating the total scattering cross section to the forward-scattering amplitude which holds in all dimensions. The procedure is to use Gegenbauer's expansion[1] of a plane wave in N dimensions in terms of partial waves appropriate for the space under consideration. This allows us to treat scattering in all dimensions on equal footing, without resort to special formulations, and reveals the similarities of wave motion in all dimensions.

The stationary-state Schrödinger equation is

$$[\Delta + k^2 - U(x)]\psi = 0, \qquad (1)$$

where Δ is the Laplacian operator in N dimensions, $k^2 = (2m/\hbar^2)E$, E is the energy, m is the mass of the particle, and $U(x) = (2m/\hbar^2)V(x)$, where $V(x)$ is the potential and x is the length of the N-dimensional position vector $\mathbf{x}$. It is thus assumed that the potential is spherically symmetric. We are interested in the scattering of a wave incident along the z axis, and ψ will depend on x and θ where $z = x\cos\theta$. The Laplacian takes the form

$$\Delta = \frac{\partial^2}{\partial x^2} + \frac{N-1}{x}\frac{\partial}{\partial x} + \frac{N-2}{x^2}\cot\theta\frac{\partial}{\partial\theta} + \frac{1}{x^2}\frac{\partial^2}{\partial\theta^2}. \qquad (2)$$

The wave equation (1) is separable, and the solutions are partial waves of the form $f_n(x)C_n^{\alpha}(\cos\theta)$, where $f_n(x)$ is the radial solution and $C_n^{\alpha}(\cos\theta)$ is the N-dimensional spherical harmonic of Gegenbauer, in which n is the degree of the polynomial and $\alpha = \frac{1}{2}N - 1$.

These harmonics are defined by the generating function[2,3]

$$(1 - 2h\cos\theta + h^2)^{-\alpha} = \sum_{n=0}^{\infty} h^n C_n^{\alpha}(\cos\theta) \qquad (3)$$

and satisfy the differential equation

$$\left(\frac{d^2}{d\theta^2} + (N-2)\cot\theta\frac{d}{d\theta} + n(N+n-2)\right)C_n^{\alpha}(\cos\theta) = 0. \qquad (4)$$

By substituting (4) and (2) into (1) we obtain the radial equation

$$\left(\frac{d^2}{dx^2} + \frac{2\alpha+1}{x}\frac{d}{dx} + k^2 - n(n+2\alpha)x^{-2} - U(x)\right)f_n(x) = 0. \qquad (5)$$

The substitution $f_n(x) = x^{-\alpha}\phi_n(x)$ gives

$$\left(\frac{d^2}{dx^2} + \frac{1}{x}\frac{d}{dx} + k^2 - (n+\alpha)^2x^{-2} - U(x)\right)\phi_n(x) = 0. \tag{6}$$

The asymptotic solution of (6) is

$$\phi_n(x) \sim Z_{n+\alpha}(kx), \tag{7}$$

where Z is a general Bessel function.

The solution ψ which is of interest must consist of an incident wave $\exp(ikx\cos\theta)$ and a scattered wave. Now, a plane wave can be resolved into partial waves by the Gegenbauer expansion formula:

$$e^{ikx\cos\theta} = 2^\alpha\Gamma(\alpha)\sum_{n=0}^{\infty}(\alpha+n)i^n(kx)^{-\alpha}J_{n+\alpha}(kx)C_n^\alpha(\cos\theta), \tag{8}$$

and, since ψ must have the asymptotic behavior of a plane wave plus outgoing partial waves, we can write the asymptotic formula

$$\psi \sim 2^\alpha\Gamma(\alpha)\sum_{n=0}^{\infty}(\alpha+n)i^n(kx)^{-\alpha}C_n^\alpha(\cos\theta) \times \tfrac{1}{2}\{e^{i\delta_n}h_{n+\alpha}^{(1)}(kx) + e^{-i\delta_n}h_{n+\alpha}^{(2)}(kx)\}e^{i\delta_n}, \tag{9}$$

where δ_n is the phase shift for the nth partial wave and $h_\nu^{(1)}$ and $h_\nu^{(2)}$ are Hankel's functions of the first and second kind. In writing (9),the well-known asymptotic behavior of Bessel functions was used, namely,[4]

$$\begin{aligned} J_\nu(x) &\sim (2/\pi x)^{\frac{1}{2}}\cos(x - \tfrac{1}{2}\nu\pi - \tfrac{1}{4}\pi), \\ h_\nu^{(1)}(x) &\sim (2/\pi x)^{\frac{1}{2}}e^{i(x-\frac{1}{2}\nu\pi-\frac{1}{4}\pi)}, \\ h_\nu^{(2)}(x) &\sim (2/\pi x)^{\frac{1}{2}}e^{-i(x-\frac{1}{2}\nu\pi-\frac{1}{4}\pi)}. \end{aligned} \tag{10}$$

By subtracting (8) from (9) we obtain the asymptotic behavior of the scattered wave ψ_s, namely,

$$\begin{aligned} \psi_s &\sim 2^\alpha\Gamma(\alpha)\sum_{n=0}^{\infty}(\alpha+n)i^n(kx)^{-\alpha}(2/\pi kx)^{\frac{1}{2}} \\ &\quad \times C_n^\alpha(\cos\theta)e^{i(kx-\frac{1}{2}(n+\alpha)\pi-\frac{1}{4}\pi)}\tfrac{1}{2}(e^{2i\delta_n} - 1) \\ &\equiv i(ix)^{-(\alpha+\frac{1}{2})}e^{ikx}f(\theta), \end{aligned} \tag{11}$$

which also defines the scattering amplitude $f(\theta)$.

The optical theorem relates the total cross section $\sigma = \int|f(\theta)|^2\,d\Omega$ to the forward scattering amplitude $f(0)$. The cross section σ can be obtained from (11) by remembering that[3]

$$\int_0^\pi \sin^{2\alpha}\theta C_m^\alpha(\cos\theta)C_n^\alpha(\cos\theta)\,d\theta = \frac{\pi\Gamma(2\alpha+n)}{2^{2\alpha-1}(\alpha+n)n!\,[\Gamma(\alpha)]^2}\delta_{m,n} \tag{12}$$

and that in the integration over solid angle the contribution of the azimuthal angles $(\phi_1, \phi_2, \cdots, \phi_{N-2})$ gives the factor Ω_ϕ, where

$$\begin{aligned} \Omega_\phi &= 2\pi^{\alpha+\frac{1}{2}}/\Gamma(\alpha+\tfrac{1}{2}), \quad \alpha \geq \tfrac{1}{2}, \\ &= 1, \quad \alpha < \tfrac{1}{2}, \end{aligned} \tag{13}$$

and thus we obtain

$$\sigma = \sum_{n=0}^{\infty}\frac{8\pi^{\alpha+\frac{1}{2}}(\alpha+n)\Gamma(2\alpha+n)}{n!\,k^{2\alpha+1}\Gamma(\alpha+\frac{1}{2})}\sin^2\delta_n. \tag{14}$$

For $\theta = 0$, Eq. (3) gives

$$C_n^\alpha(1) = (n!)^{-1}\Gamma(2\alpha+n)/\Gamma(2\alpha), \tag{15}$$

which enables us to deduce from Eq. (11) that

$$\operatorname{Im}f(0) = 2^{\alpha+\frac{1}{2}}\pi^{-\frac{1}{2}}\frac{\Gamma(\alpha)}{\Gamma(2\alpha)}k^{-(\alpha+\frac{1}{2})} \times \sum_{n=0}^{\infty}\frac{(\alpha+n)\Gamma(2\alpha+n)(\sin^2\delta_n)}{n!}. \tag{16}$$

By comparing (14) and (16) we obtain the optical theorem:

$$\sigma = 8\pi^{\alpha+1}(2k)^{-(\alpha+\frac{1}{2})}[\Gamma(2\alpha)/\Gamma(\alpha)][\Gamma(\alpha+\tfrac{1}{2})]^{-1}[\operatorname{Im}f(0)]. \tag{17}$$

In three dimensions, $\alpha = \frac{1}{2}$, and we have the well-known result $\sigma = (4\pi/k)\operatorname{Im}f(0)$. In two dimensions $\alpha = 0$ and

$$\sigma(N=2) = 2(2\pi/k)^{\frac{1}{2}}[\operatorname{Im}f(0)]. \tag{18}$$

In one dimension $\alpha = -\frac{1}{2}$ and

$$\begin{aligned} \lim_{\alpha\to-\frac{1}{2}}\frac{\Gamma(2\alpha)}{\Gamma(\alpha)\Gamma(\alpha+\frac{1}{2})} &= \lim_{\alpha\to-\frac{1}{2}}\frac{2\alpha\Gamma(2\alpha)(\alpha+\frac{1}{2})}{2\alpha\Gamma(\alpha)(\alpha+\frac{1}{2})\Gamma(\alpha+\frac{1}{2})} \\ &= \tfrac{1}{4}[\Gamma(2\alpha+2)/\Gamma(\alpha+1)][\Gamma(\alpha+\tfrac{3}{2})]^{-1} \\ &= \tfrac{1}{4}\pi^{-\frac{1}{2}} \end{aligned}$$

(which also could have been foreseen from the duplication formula for the gamma function), and thus we obtain

$$\sigma = 2\operatorname{Im}f(0), \quad N = 1, \tag{19}$$

a result which is equivalent to that obtained by Eberly,[5] since our $f(0)$ as defined by Eq. (11) is $-i$ times the $f(0)$ used by Eberly.

Other applications of the Gegenbauer expansion formula have been given elsewhere.[6]

[1] G. N. Watson, *A Treatise on the Theory of Bessel Functions* (Cambridge U.P., Cambridge, 1966).
[2] Reference 1, p. 129.
[3] A. Sommerfeld, *Partial Differential Equations in Physics* (Academic, New York, 1949), Appendix IV.
[4] Reference 1, p. 198.
[5] J. H. Eberly, Am. J. Phys. **33**, 771 (1965).
[6] I. Adawi, Phys. Rev. **146**, 379 (1966); Phys. Letters **26A**, 317 (1968).

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 12, NUMBER 3 MARCH 1971

Some Aspects of the Relationship between Mathematical Logic and Physics. II*

PAUL A. BENIOFF
Argonne National Laboratory, Argonne, Illinois 60439

(Received 24 July 1970)

In this work, a definition of agreement between a physical theory and experiment, proposed in earlier work, is extended to be relative to τ where τ is Zermelo–Fraenkel set theory. The main aim of this work is to show that this definition, unlike that of earlier work, is sufficiently powerful to include relations between limit properties of empirical outcome sequences and expectation values obtained from the physical theory. We also extend, to the more powerful τ, some earlier results on randomness and the empirical determinability of the probability measure which a physical theory assigns to the outcome set of an infinite sequence of experiments.

I. INTRODUCTION

In this paper, the program begun in an earlier paper,[1] hereafter referred to as I, is continued. Basically the aim of the program is to understand and describe the dynamics of the knowledge acquisition process, the importance of which to physics has been stressed by several authors.[2–4] If one believes, as this author does, that physical reality, far from being independent of epistemology, is, in essence, defined by the basic aspects of the process, then a study of this process should aid in understanding some basic aspects of physics (and mathematics).

The goal of the epistemological process which is of relevance here is the construction of a comprehensive physical theory as an interpreted formal axiomatized system which is valid or "explains the physical world." An understanding of the process clearly requires that one have some reasonably exact idea of what it means for a physical theory to agree with experiment. It is with this problem that I and this paper are mainly concerned.

In I an exact definition of this concept as a necessary condition of validity for a large class of physical theories was proposed and discussed. A physical theory was considered to correspond to a mapping U with domain D_U in the set of all infinite sequences Qst, of instructions for carrying out infinite sequences of single measurements, and with range in the class of probability measures defined on the usual σ-algebra $A(\Omega)$ of subsets of Ω. The set of all infinite sequences of natural numbers is Ω. In essence the definition stated that, for each (Qst) in the domain of a physical theory, there is a (in general Qst-dependent) class of properties which ψ_{Qst}, the infinite outcome sequence of natural numbers obtained by actually carrying out (Qst), must have if the theory is to agree with experiment. For each (Qst) the class of properties was taken to be a subclass of all those properties of elements of Ω which are expressible by some formula in H. H was taken to be a certain extension of the formulas of the formal language of elementary analysis or second-order arithmetic.[5]

In I it was seen that, although this definition included much of the intuitive meaning of agreement between a physical theory and experiment, there were some important aspects left out. In particular, the definition did not include any relation between the limit mean of an empirical outcome sequence and an expectation value given by the physical theory. It was noted that the reason for this was that H was not strong enough to include the necessary probability theoretic statements.

The main purpose of this work is to remedy this defect. In essence this is done by replacing H in the definitions of I by τ, where τ is basic Zermelo–Fraenkel set theory.[6–8] The change to set theory requires some changes in the definitions and procedures over those used in I, but does not appear to cause any difficulty.

Section V contains the basic results of this work. In particular, Metatheorem 6 states that if a physical theory U τ-agrees with experiment [Eqs. (12) and (38)], then for each (Qst) in D_U there are different conditions which the measure $U(Qst)$ must satisfy if different relations between τ-definable limit properties of sequences and general expectations computed from theory are to hold for the outcome sequence ψ_{Qst}. The essence of the metatheorem is an ergodic theorem and related aspects, as the conditions on $U(Qst)$ are given in terms of the various ergodic, etc., properties of τ-definable transformations. For example, a particular case covered by the metatheorem is the following: If U τ-agrees with experiment, then for any (Qst) in D_U, if the one-sided shift operator Eq. (19) is ergodic and measure preserving with respect to the measure $U(Qst)$, $\bar{M}\psi_{Qst} = \sum_n nU(Qst)E_{n0}$. $\bar{M}\psi_{Qst}$ is the limit mean of ψ_{Qst} and

$$E_{n0} = [\varphi \mid \varphi(0) = n].$$

This metatheorem (and Metatheorems 4 and 5) show that Eqs. (12) and (38) include those relations between limit properties of sequences and theoretical expectations which were left out of the definition given before.[1] Thus, at least for the type of theories discussed here, Eqs. (12) and (38), with τ equal to ZF set theory, appear to carry the full intuitive meaning of agreement between theory and experiment.

Metatheorems 7 and 8 extend the corresponding results of I. Metatheorem 7 states that if U τ-agrees with experiment, then, for all (Qst) in the domain of U, if $U(Qst)$ is a product measure, the outcome sequence obtained by doing Qst is τ-random [Eq. (14)]. Metatheorem 8 gives sufficient conditions for the measure $U(Qst)$ to be empirically determinable. The concept of empirical determinability of $U(Qst)$ is defined by Eq. (53).

Sections II–IV give the mathematical background necessary for an understanding of Sec. V. Section II reviews the formal linguistic framework by first giving a standard axiomatization of ZF set theory.[6,7] The method of extension of a formal theory by definitions,[6] which gets much implicit use, is reviewed. If desired, this section can be skipped on first reading as the material, although fundamental, is, for the most part, used implicitly only.

Section III extends the definition given in I for the statement $E(\tau, P, \varphi)$ [Eq. (12)]. This statement, which is central to the whole work, says in effect that all τ-definable properties of sequences (where the probability measure P may occur in the defining relation) which hold for P almost all sequences hold for the sequence φ. Metatheorems 1–3 give some properties of this definition. Metatheorem 1 says the definition is not empty; for any P there exist sequences such that $E(\tau, P, \varphi)$ holds. Metatheorem 3 says that for all nontrivial measures P if $E(\tau, P, \varphi)$ holds, then φ is not τ-definable. Also the definition of randomness given in I is extended to the definition of $E(\tau, P, \varphi)$ given here.

Section IV develops the relevant probability theoretic background and gives the metatheorems necessary to prove Metatheorem 6. First the standard ergodic and indecomposability theorems of probability theory are extended to include transformations from one measure space to another. Then it is shown (Metatheorem 4) that statements which relate limit properties of sequences to expectations correspond to formulas in τ. Furthermore, by the ergodic, etc., theorems, such statements, for τ-definable random variables and transformations, are, under certain conditions, true P almost everywhere. Thus if $E(\tau, P, \psi)$ holds, they must be true for ψ. This is the content of Metatheorem 5.

Finally, some general comments are in order. In this work the distinction is made between theorems as statements formalizable and provable within set theory and metatheorems which include semantic or metalinguistic aspects outside of formal set theory. Considerations of whether or not the metatheorems can be converted into theorems of τ are outside the scope of this work.

As in I, the position taken here is that we are outside of any particular physical theory and are considering some aspects of the general process of constructing a valid physical theory. This is one reason why the discussion is not carried out in quantum mechanics. Furthermore, as was discussed in I, it is an open question whether or not quantum mechanics is a theory of the type U as discussed here. In particular, there seem to be many problems in attempting to discuss correlations between single measurements within quantum mechanics. Any theory U which, for each appropriate (Qst), defines a measure $U(Qst)$ on $A(\Omega)$ is a theory which automatically describes correlations between single measurements.[1]

II. FORMAL LINGUISTIC FRAMEWORK

The formal linguistic framework for this work is that of Zermelo–Fraenkel set theory, here denoted by τ. The symbols of the language, $L(\tau)$, of τ consist of $\in$ (membership) as the one primitive predicate symbol, $x, y, z, \cdots$ as set variables, and the logical symbols $\vee$ (or), $\wedge$ (and), $\neg$ (not), $\Rightarrow$ (implies), $\exists$ (there exists), $\forall$ (for all), and $=$ as the equality predicate symbol.

The formulas of $L(\tau)$ are the smallest class of expressions closed under the logical operations which contain the atomic formulas. That is, (1) $u \in v$ and $u = v$ are formulas where u and v are any set variables, and (2) $\neg B$, $B \vee B'$, $B \wedge B'$, $B \Rightarrow B'$, $\exists u B$, and $\forall u B$ are formulas if B and B' are formulas and u is any set variable. The list of logical connectives used here is convenient even though there is redundancy. For example, $B \wedge B' \equiv \neg(\neg B \vee \neg B')$, $B \Rightarrow B' \equiv \neg B \vee B'$, and $\forall u B \equiv \neg \exists u \neg B$. In what follows,free use is made of parentheses and brackets in order to increase the readability of the formulas. They are not necessary, though, and can be removed by an appropriate reordering of the symbols in the formulas.

There are several equivalent axiomatizations of basic ZF set theory. We give one of them[6] here.

(1) The axiom of extensionality:

$$\forall x \forall y \forall z\, ((z \in x \Leftrightarrow z \in y) \Rightarrow x = y).$$

This says two sets are equal if they have the same members ($A \Leftrightarrow B$ is short for $A \Rightarrow B \wedge B \Rightarrow A$). This axiom also says that a set is determined by its members.

(2) The regularity axiom:

$$\forall x\,[\exists y\,(y \in x) \Rightarrow \exists y\,(y \in x \wedge \forall z\,(z \in x \Rightarrow \neg z \in y))].$$

This says that if x is not empty, then it contains a minimal element. That is, it contains an element y such that x and y have no members in common.

(3) The subset axiom schema:

$$\forall yz\exists\forall x\,(x \in z \Leftrightarrow x \in y \wedge A).$$

This is really a statement which stands for an infinite number of axioms, one for each formula A of $L(\tau)$ which does not contain y and z free. This says that, for each formula A and each set y, there is a set z which is the set of all and only those elements of y for which A holds.

(4) The replacement axiom schema:

$$[\forall x\exists z\forall y\,(A \Leftrightarrow y \in z)] \Rightarrow \forall w\exists z\forall y\,[\exists x\,(x \in w \wedge A) \Rightarrow y \in z].$$

Like the subset axiom schema, this is equivalent to an infinite number of axioms, one for each formula A which does not contain z and w free. To see what this means, let A be a formula with x and y free. Then this says that if there exists a set z_x containing all and only those elements y for which $A(x, y)$ holds, then for each w there exists a set z_w containing all elements y such that $A(x, y)$ holds for some x in w.

(5) The power set axiom:

$$\forall x\exists y\forall z\,(z \subseteq x \Rightarrow z \in y).$$

This says that for any set x there is a set y (called the power set of x) containing as elements all subsets of x. Here $z \subseteq x$ is short for $\forall w\,(w \in z \Rightarrow w \in x)$.

(6) The axiom of infinity:

$$\exists x\,\{\varnothing \in x \wedge \forall y\,[y \in x \Rightarrow \exists w\,(\forall z\,(z \in w \Leftrightarrow z \in y \vee z = y) \wedge w \in x)]\}.$$

This says that there exists a set x containing an infinite number of elements. That is, it contains the empty set $\varnothing$, and for all y, if it contains y, it contains the successor set of y. $\varnothing$ is defined by $\exists x\forall y\,(\neg y \in x)$.

(7) The axiom of choice[7]:

$$\forall x\,[\forall y\forall z\,[y \in x \wedge z \in x \wedge \neg(z = y) \wedge \neg(y = \varnothing) \Rightarrow y \cap z = \varnothing] \Rightarrow \exists w\forall y]y \in x \Rightarrow \exists z\,(z \in y \Leftrightarrow z \in w) \wedge \forall v\forall u\,(u \in y \wedge u \in w \wedge v \in y \wedge v \in w \Rightarrow u = v).$$

This axiom says that for each set x whose members are pairwise disjoint, there exists a set w containing exactly one element from each member of x. Here $y \cap z$ denotes the set theoretic intersection of x and y.

Besides these nonlogical axioms there are the logical axioms and rules of deduction. These are common to every formal system and can be given in any one of several equivalent forms.[6] The logical axioms include $B \vee \neg B$ for any formula B of $L(\tau)$, $u = u$ for any variable u, etc. In one form, the rules of deduction include: If A is a theorem and $A \Rightarrow B$ is a theorem, then B is a theorem.

One now defines a theorem of set theory as a provable formula. Equivalently, the set of theorems is the smallest set of formulas of $L(\tau)$ which contains all the axioms and is closed under the logical deduction operations.

The axioms as presented contain some defined symbols such as $\varnothing$, $\subseteq$, etc., which are not in the original language. This is an example of the extension of $L(\tau)$ and τ by definitions.[6] Since this process is given much implicit use in the following, it is worthwhile to give a brief review.

An unary predicate symbol R is added to $L(\tau)$ along with

$$\forall x\,[R(x) \Leftrightarrow A(x)] \tag{1}$$

as the defining axiom for R, where $A(x)$ is a formula of $L(\tau)$ which defines $R(x)$.

A symbol S is added as the name of a set along with the defining relation

$$\forall x\,(x \in S \Leftrightarrow A(x)) \tag{2}$$

as the defining axiom for S provided that

$$\exists z\forall x\,(x \in z \Leftrightarrow A(x)) \tag{3}$$

is a theorem of τ.

An unary function symbol f is added along with

$$\forall x\forall y\,[f(x) = y \Leftrightarrow A(x, y)] \tag{4}$$

as a defining relation with $A(x, y)$ a formula of $L(\tau)$ provided that

$$\forall x\forall y\,[A(x, y) \wedge A(x, y') \Rightarrow y = y'], \tag{5}$$

$$\exists x\exists y A(x, y), \tag{6}$$

$$\exists z\forall x\,[x \in z \Leftrightarrow \exists y A(x, y)] \wedge \exists z\forall y\,[y \in z \Leftrightarrow \exists x A(x, y)] \tag{7}$$

are theorems of τ. Conditions (5) and (6) ensure that f is a function with a nonempty domain, and condition (7) ensures that f, as a class of ordered pairs, is a set by requiring that the domain and range of f be sets. These defining conditions have obvious extensions to n-ary relation and function symbols.

As an example, suppose one wants to add to $L(\tau)$ the formula $\underline{N}(x)$, which says that x is a natural number. One then adds the predicate symbol $\underline{N}$ to $L(\tau)$ along with the defining axiom

$$\underline{N}(x) \Leftrightarrow \text{Tran}(x) \wedge \forall y\,(y \in x \Rightarrow \text{Tran}(y) \wedge \neg \text{Inf}(x)), \tag{8}$$

where

$$\text{Tran}(x) \equiv \forall y \forall z\,(y \in x \wedge z \in y \Rightarrow z \in x) \tag{9}$$

and $\text{Inf}(x)$ is the axiom of infinity.

One can further add $\tilde{N}$ as a symbol for the set N of all natural numbers to τ, along with the defining axiom

$$x\,(x \in \tilde{N} \Leftrightarrow \underline{N}(x)), \tag{10}$$

provided that $\exists z \forall x\,(x \in z \Leftrightarrow \underline{N}(x))$ is a theorem of τ. The proof of this is carried out by a standard construction[6] in which one shows that there exists a set which is a limit ordinal and that N (or ω) is the first limit ordinal.

The reason the existence proofs, expressions (3) and (7), are necessary is that many formulas of τ define classes which are not sets. Thus one must prove that the class of sets with the property defined by the formula in question is a set before a name of a class can be admitted as a name of a set.

These extensions by definitions do not make any essential change in $L(\tau)$ or τ. This is shown by the fact that if τ' is an extension of τ by definitions, then[6] for each formula A' in the extended language $L(\tau')$, A' is a theorem of τ' if and only if A is a theorem of τ where A is a translation of A' back to the original language $L(\tau)$. Of course, $A \equiv A'$ if A' contains no defined symbols.

It is important to keep in mind that formulas of $L(\tau)$ are meaningless strings of symbols which are given meaning by interpretation in a model of τ. A model of τ is a structure for $L(\tau)$ in which all the axioms are valid. A structure for $L(\tau)$ is a universe of elements Z, together with functions and relations defined on the universe and which serve as interpretations of the function and relation symbols in $L(\tau)$. Similarly, the variables and constants, or names of sets, are mapped into Z. In any structure, each set name has a fixed interpretation whereas the meaning of a variable can range over all elements of Z. A formula is valid in a structure if it has no free variables and is true, or if it is true for all interpretations of the free variables which it may have.

From now on definitions of various mathematical objects will be introduced without any reference to their formal definitions in τ. This is the procedure one follows when one builds up mathematics in informal set theory,[6-8] which is the standard model of τ. Of course, one must be careful to ensure that every informal definition can be replaced by a formal definition in $L(\tau)$. Thus, if desired, one can introduce, by the extension process just described, a name and defining axiom for the defined object into $L(\tau)$.

Furthermore, the formal logical and relation symbols of $L(\tau)$ will also be used to denote their informal, meaningful counterparts. For example, $\in$ and $=$ will also be used to denote the respective meaningful membership and equality relations. This identification of formal objects of $L(\tau)$ and informal objects should cause no confusion since it will always be clear from context whether the discussion is about informal objects or formal objects of τ.

III. $E(\tau, P, \psi)$

Some definitions are now in order. Let N denote the set of natural numbers [with name $\tilde{N}$, Eq. (10)] and $i, j, k, i', \cdots$ denote informal variables which range over N. The natural numbers of N are given by $0, 1, \cdots, n, l, m, \cdots$, with $0 = \varnothing$, $1 = \{\varnothing\}$, $2 = \{\varnothing, \{\varnothing\}\}$, etc., with $\varnothing$ denoting the empty set. Let Ω be the set of all infinite sequences of natural numbers and $\varphi, \theta, \psi, \cdots$ be informal variables for such sequences. Let $E, F, \cdots$ be informal variables ranging over all subsets of Ω. For each n and l, E_{nl} is the subset of Ω defined by

$$\forall \varphi\,(\varphi \in E_{nl} \Leftrightarrow \varphi(l) = n). \tag{11}$$

(Equivalently one can write $E_{nl} \equiv [\varphi \mid \varphi(l) = n]$.)

Let R denote the set of real numbers. The operations $+$ (plus) and $\times$ (times) are defined on N, R, and Ω in the usual manner.[8] Thus $(\varphi + \theta)(j) = \varphi(j) + \theta(j)$ and $(\varphi \times \theta)(j) = \varphi(j) \times \theta(j)$. Let $A(\Omega)$ be the minimal σ-field of subsets of Ω generated by the E_{nl} and Pr the set of probability measures defined on $A(\Omega)$, with $P, P', \cdots$ denoting informal variables ranging over Pr.

The relation $E(\tau, P, \psi)$, which is central to this work, can now be defined as follows:

$$\begin{aligned} E(\tau, P, \psi) \equiv \forall q\,[&q \text{ has exactly one free sequence} \\ &\text{variable and at most one free probability measure} \\ &\text{variable and } [\varphi \mid q_P(P, \varphi)] \in A(\Omega) \text{ and} \\ &P[\varphi \mid q_P(P, \varphi)] = 1 \Rightarrow q_P(P, \psi)]. \end{aligned} \tag{12}$$

Here q is a syntactic variable which ranges over all formulas in $L(\tau)$, the language of Zermelo-Fraenkel set theory, τ. A variable is bound in a formula if it is acted on by either of the quantifiers $\exists$ or $\forall$; otherwise it is free. The formula q_P is obtained from q by setting

$q \equiv q_P$ if q already has P free or has no free probability measure variables. If q has a probability measure variable other than P free, then one replaces all free occurrences of the variable by P. If P is already bound in q, one must first change the bound P to a different variable (e.g., P'') before the replacement is made. This is a standard procedure for changing the variables of an expression without changing its meaning.

In words $E(\tau, P, \psi)$ is the statement that, for all formulas q of $L(\tau)$ which contain exactly one free sequence variable and at most one free probability measure variable, if q is such that the set of all φ for which $q_P(P, \varphi)$ is true is in $A(\Omega)$ and is a set of P measure 1 (or is true P almost everywhere), then $q_P(P, -)$ is true at ψ. Equivalently, $E(\tau, P, \psi)$ says that ψ is contained in the intersection of all τ-definable subsets of Ω which are elements of $A(\Omega)$ and are sets of P measure 1, where P can occur in the defining relation for the subsets.

It should be noted that the semantic concept of truth is not an essential part of the definition of $E(\tau, P, \psi)$. The reason is that one can replace

$$[\varphi \mid q_P(P, \varphi)] \in A(\Omega)$$

and $P[\varphi \mid q_P(P, \varphi)] = 1$ in the definition by the equivalent expression

$$\exists E\,[\forall \varphi\,(\varphi \in \mathsf{E} \Leftrightarrow q_P(P, \varphi)) \wedge E \in A(\Omega) \wedge PE = 1],$$

which (see below and Sec. II) can be converted to a formula of τ. In a strict sense, $q_P(P, \varphi)$ is not a formula of τ, as P and φ are informal elements and not symbols of $L(\tau)$. However, $q_P(P, \varphi)$ corresponds to the formula $q_x(x, y) \wedge \mathrm{Pr}\,(x) \wedge \mathrm{seq}\,(y)$ in $L(\tau)$, where $\mathrm{Pr}\,(x)$ and $\mathrm{seq}\,(y)$ are the defining relations in τ for "x is a probability measure on $A(\Omega)$" and "y is an infinite sequence of natural numbers." Similar arguments apply to expressions containing E, F, i, j, etc.

Some useful properties of $E(\tau, P, \psi)$ are given by the following three metatheorems. The proofs will not be given, as they are given elsewhere.[9]

The first metatheorem says that the definition of $E(\tau, P, \psi)$ is not empty.

Metatheorem 1 (Existence): For all P there exist ψ such that $E(\tau, P, \psi)$ holds. In fact there exist many sequences ψ such that $E(\tau, P, \psi)$ holds, since the proof[9] consists of showing that $P[\psi \mid E(\tau, P, \psi)] = 1$.

Metatheorem 2: Let q be any formula of τ with one free sequence variable and at most one free probability measure variable. Then for any φ and P if $E(\tau, P, \varphi)$ holds and $[\varphi \mid q_P(P, \varphi)] \in A(\Omega)$, then

$$P[\varphi \mid q_P(P, \varphi)] = 0 \Rightarrow \neg q_P(P, \psi).$$

Some definitions are needed for the next metatheorem. A probability measure P is nontrivial if

$$\forall \varphi\,(P\{\varphi\} = 0). \tag{13}$$

This excludes pathological measures which are nonzero on any single element subset of Ω.

A formula q in $L(\tau)$ (implicitly) defines a sequence in Ω if q includes the defining conditions for sequences in Ω, has exactly one free variable, and is such that

$$\exists \varphi\, q(\varphi) \wedge \forall \varphi \forall \psi\,[q(\varphi) \wedge q(\psi) \Rightarrow \varphi = \psi]$$

is a theorem in τ.

Metatheorem 3 (Undefinability): For any sequence ψ, if there exists a nontrivial measure P such that $E(\tau, P, \psi)$ holds, then for all formulas q in $L(\tau)$, if q defines a sequence in Ω, we have $\neg q(\psi)$.

The proof[9] consists of showing that if q defines a sequence in Ω, then $[\varphi \mid \neg q(\varphi)]$ is the complement of a one-point set and is thus a set y, for q, such that $y \in A(\Omega)$ and $Py = 1$ for P nontrivial. Thus, by Eq. (12), we have $\neg q(\psi)$.

This theorem places strong limitations on which sequences are definable. To see this, one notes that an equivalent statement of this theorem is that, for all ψ, if there exists a nontrivial P such that $E(\tau, P, \psi)$ is true, then ψ is not definable in τ. Furthermore, this holds in any model of τ. Conversely, any τ-definable sequence ψ is such that $E(\tau, P, \psi)$ is false for all nontrivial P. Also, this and the existence metatheorems place strong limitations on the arbitrary introduction of names of elements of Ω or R into $L(\tau)$.[9]

The definition of τ-randomness is an extension of that given before.[1]

$$R_\tau(q) \equiv \exists P\,[P \text{ is a nontrivial product probability measure defined on } A(\Omega) \text{ and } E(\tau, P, q)]. \tag{14}$$

This definition is similar to that given by Kruse[10] and includes all the tests for randomness given in the other literature definitions.[11–13] In particular, the definitions which consider subsequence selection procedure tests are included.[12,13]

It is also clear from the definition and the existence metatheorem that there exist τ-random sequences. In fact, the set of τ-random sequences coincides with the set $\bigcup_P [\varphi \mid E(\tau, P, \varphi)]$, where the union is over all nontrivial product measures. Also, from the undefinability metatheorem, one has the result that any τ-random sequence is not τ-definable. This can be seen by the fact that all sequences in $\bigcup_P [\varphi \mid E(\tau, P, \varphi)]$,

where the union is over all nontrivial measures, are not definable in τ.

IV. RELEVANT PROBABILITY THEORY ASPECTS

The aim of this section is to develop from the relevant probability theory background the result that many statements, which relate limit properties of sequences to appropriate expectation values, are formulas $q(P, \varphi)$ of τ. Furthermore, conditions are given under which many of these statements hold P almost everywhere. Thus, by the definition of $E(\tau, P, \psi)$, if these conditions are satisfied, the statements hold at ψ. Ergodic and indecomposibility theorems are central to the argument.

A. Background

Some definitions are in order. Let Ω, $A(\Omega)$, and P be defined as before. Let O be a point transformation with domain D_O and range R_O both subsets of Ω such that $D_O \in A(\Omega)$. Any such O induces a set transformation O' such that for each $B \subseteq R_O$ and $O'B \subseteq D_O$, with $O'B$ defined by ($B \subseteq D$ means B is a subset of D),

$$O'B = [\varphi \mid O\varphi \in B]. \tag{15}$$

Define $A(D_O)$ and $A(R_O)$ to be the restrictions of $A(\Omega)$ to D_O and R_O, respectively. That is, for each $E \in A(\Omega)$ define E_D and E_R by

$$E_D = E \cap D_O, \tag{16}$$

$$E_R = E \cap R_O, \tag{17}$$

and let $A(D_O)$ be a collection of subsets of D_O such that, for each set in the collection, Eq. (16) holds for some E in $A(\Omega)$. $A(R_O)$ is defined similarly from Eq. (17). One sees from these definitions that the mappings $E \to E_D$ and $E \to E_R$ are many one-mappings from $A(\Omega)$ to $A(D_O)$ and $A(R_O)$ which preserve all countable set operations. Thus $A(D_O)$ and $A(R_O)$ are σ-fields. O is a measurable transformation[14] if, for each $E_R \in A(R_O)$, $O'E_R \in A(D_O)$.

The reason such transformations must be included is that the discussion must be able to handle subsequence selection procedures. A subsequence selection procedure is a mapping O defined for all j and $\varphi \in D_O$ by

$$(O\varphi)(j) = \varphi(g(\varphi, j)). \tag{18}$$

D_O is the set of all φ such that $g(\varphi, j)$ is defined for all j where D_O can be a proper subset of Ω. For example, in the selection procedure "select, in the natural order, each element which follows the occurrence of a 1 by one place"; O is not defined for any sequence having only a finite number of ones.

Let T denote any point transformation with domain D_T and range R_T equal to Ω and which is measurable, and let T' be the corresponding set transformation, Eq. (15). Define a set transformation to be structure preserving on Ω if it preserves all countable set operations and takes Ω to Ω and $\varnothing$ to $\varnothing$.[15] It is easy to see from Eq. (15) that any such T' is structure preserving on Ω. Clearly, any T is a special case of the definition of O. An example which will be much used later on is the one-sided shift operator defined for all φ and j by

$$(T\varphi)(j) = \varphi(j + 1). \tag{19}$$

For any O let T_O and $T_{O'}$ denote the restriction of T to D_O and T' to R_O. That is, for each $B \subseteq R_O$,

$$T_{O'}B = T'B. \tag{20}$$

One has the following theorem.

Theorem 1: Let O be any point transformation with $D_O \subseteq \Omega$ and $R_O \subseteq \Omega$ with $A(R_O)$ as defined by Eq. (17). Let T be a measurable point transformation with domain and range equal to Ω, and let $T_{O'}$ be the restriction of T to R_O. Then, if O and T are such that $T'R_O = R_O$, $T_{O'}$ is structure preserving on R_O and maps $A(R_O)$ into $A(R_O)$.

Proof: The proof that $T_{O'}$ is structure preserving on R_O and maps $A(R_O)$ into $A(R_O)$ proceeds by transfinite induction on the generating process for $A(R_O)$. By Eq. (17) and the definition of $A(\Omega)$, $A(R_O)$ is a system of subsets of R_O closed under R_O-complementation and the formation of countable unions and which contains the E_{nlR} for $n, l = 0, 1, \cdots$ [Eqs. (11) and (17)].

Since T' is measurable and structure preserving on Ω and $T'R_O = R_O$, one has, for each n, l,

$$\begin{aligned} T_{O'}E_{nlR} &= T'(E_{nl} \cap R_O) \\ &= T'E_{nl} \cap R_O \in A(R_O) \quad \text{since} \quad T'E_{nl} \in A(\Omega). \end{aligned}$$

Suppose for any $E_R \in A(R_O)$ that $T_{O'}E_R \in A(R_O)$. Then, since $A(R_O)$ is closed under R_O-complementation, one has ($E_R^C = R_O - E_R$),

$$\begin{aligned} T_{O'}(E_R^c) &= T'(R_O - E \cap R_O) \\ &= R_O - T'E \cap R_O = (T_{O'}E_R)^c \in A(R_O). \end{aligned}$$

Finally, let E_{R_i} and $T_O E_{R_i}$ all be in $A(R_O)$ for $i = 0, 1, \cdots$. Then, since $A(R_O)$ is closed under the formation of countable unions,

$$\begin{aligned} T_{O'}\left(\bigcup_i E_{R_i}\right) &= T' \bigcup_i (E_i \cap R_O) = \bigcup_i (T'E_i \cap R_O) \\ &= \bigcup_i T_{O'}E_{R_i} \in A(R_O). \end{aligned}$$

QED

It is to be noted that, by referring to this and the following as theorems and not metatheorems, it is implied that the statement of each theorem is a formula of ZF set theory and the formula can be proved within the theory. A proof of this is not given here. However, the theorems given here are clearly informal theorems of probability theory. Thus, since one can conservatively extend set theory to include the defining axioms of probability theory, the informal theorems and their proofs should be formalizable within set theory.

Let P be a probability measure defined on $A(\Omega)$, and let O be a measurable transformation such that $D_O \in A(\Omega)$ and $PD_O > 0$. Then P induces probability measures P_D and P_O defined on $A(D_O)$ and $A(R_O)$, respectively, by

$$P_D E_D = \frac{1}{PD_O} PE_D \tag{21}$$

for each $E_D \in A(D_O)$ and

$$P_O E_R = P_D O' E_R \tag{22}$$

for each $E_R \in A(R_O)$.

As further definitions let $\bar{P}_{On}$ and $\bar{P}_O$ be the probability measures[15] defined for each E_R in $A(R_O)$ by

$$\bar{P}_{On} E_R = \frac{1}{n} \sum_{j=0}^{n-1} P_O (T')^j E_R \tag{23}$$

and

$$\bar{P}_O E_R = \lim_{n \to \infty} P_{On} E_R, \tag{24}$$

if the limit exists. [To save on notation, we use Eq. (19) freely and use T' instead of $T_{O'}$.]

Finally, let X be a measurable function from R_O to R, the set of real numbers. For any T such that $T'R_O = R_O$, T induces a transformation $\mathcal{T}$ on the set of all such functions defined by

$$\mathcal{T}X(\varphi) = X(T\varphi) \tag{25}$$

for all $\varphi \in R_O$. (Since $T'R_O = R_O$, $T\varphi \in R_O$ if $\varphi \in R_O$.) For each X, define $\bar{X}_n$ and $\bar{X}$ by

$$\bar{X}_n(\varphi) = \frac{1}{n} \sum_{j=0}^{n-1} \mathcal{T}^j X(\varphi) = (\bar{\mathcal{T}}_n X)(\varphi) \tag{26}$$

and

$$\bar{X}(\varphi) = \lim_{n \to \infty} \bar{X}_n(\varphi) = (\bar{\mathcal{T}} X)(\varphi), \tag{27}$$

where $\bar{X}(\varphi)$ is defined if and only if the limit exists.

The two main theorems from probability theory which are needed here are an ergodic theorem and the indecomposibility theorem. An appropriate ergodic theorem[15] extended to include the transformations O is given below.

Theorem 2 (Ergodic): Let $T: \Omega \to \Omega$ be a measurable, structure-preserving transformation with domain and range equal to Ω and O a measurable transformation with $D_O \subseteq \Omega$ and $R_O \subseteq \Omega$ such that (1) $D_O \in A(\Omega)$ and $R_O \in A(\Omega)$, (2) $PD_O > 0$, and (3) $T'R_O = R_O$. Let T_O be the restriction of T to R_O and $\bar{P}_{On}$ be defined by Eq. (23).

If $\lim_n \bar{P}_{On} = \bar{P}_O$ exists on $A(R_O)$ and is a probability measure, then, for every nonnegative measurable function X, $\lim_n \bar{X}_n(\varphi)$ exists and

$$\lim_{n \to \infty} \bar{X}_n(\varphi) = (\bar{E}_O^{\Gamma} X)(\varphi)$$

almost everywhere.

In this theorem Γ is the sub σ-field of $A(R_O)$ of invariant subsets of R_O, i.e., $T_{O'}C = C$ for each $C \in \Gamma$ and $\bar{P}_O = P_O$ on Γ. $\bar{E}_O^{\Gamma} X$ is the Γ-measurable T_O-invariant function defined as the restriction of the indefinite integral of X to Γ by[15]

$$\int_C (\bar{E}_O^{\Gamma} X)(\varphi)\, d\bar{P}_{O\Gamma}(\varphi) = \int_C X(\varphi)\, d\bar{P}_O(\varphi) \tag{28}$$

for each $C \in \Gamma$. The existence and $\bar{P}_{O\Gamma}$ almost-everywhere uniqueness of $\bar{E}_O^{\Gamma} X$ is guaranteed by the Radon–Nikodym theorem.[14] Here $\bar{P}_{O\Gamma}$ is the restriction of $\bar{P}_O$ to Γ. The T_O invariance of $\bar{E}_O^{\Gamma} X$ means that

$$(\bar{E}_O^{\Gamma} X)(T_O \varphi) = \bar{E}_O^{\Gamma}(X)(\varphi),$$

$\bar{P}_{O\Gamma}$ almost everywhere.

If P_O, T, and Γ are such that there is a sub σ-field Γ' which is P_O equivalent to Γ and Γ' is generated by a countable partition $\{C_i\}$ with $i = 0, 1, \cdots$ of R_O, then $\bar{E}_O^{\Gamma} X$ can be given the more explicit definition[15]

$$(\bar{E}_O^{\Gamma} X)(\varphi) = \sum_i (\bar{E}_{OC_i} X) I_{C_i}(\varphi), \tag{29}$$

which holds $\bar{P}_O$ almost everywhere in R_O.

Here I_{C_i} is the characteristic function for the set C_i. For those i for which $\bar{P}_O C_i > 0$, $\bar{E}_{OC_i} X$, the conditional expectation of X, given C_i, is defined by

$$\bar{E}_{OC_i} X = \frac{1}{\bar{P}_O C_i} \int_{C_i} X(\varphi)\, d\bar{P}_O(\varphi). \tag{30}$$

If $\bar{P}_O C_i = 0$, then $\bar{E}_{OC_i} X$ is undefined, and so $(\bar{E}_O^{\Gamma} X)(\varphi)$ is undefined for any φ such that $I_{C_i}(\varphi) = 1$ and $\bar{P}_O C_i = 0$.

The proof of this theorem will not be given here as it is given elsewhere[15] and is rather long. It is to be noted that, from the conditions given in the theorem for P, T, and O, the triple $(R_O, A(R_O), P_O)$ is a probability measure space with $T_{O'}$ a translation[15] on $A(R_O)$ (Theorem 1). Thus the proof given by Loeve applies directly to $A(R_O)$-measurable functions on R_O.

The ergodic hypothesis is equivalent to the statement that $T_{O'}$ is P indecomposable, or that, up to a P_O equivalence, the σ-field Γ contains only the empty set $\varnothing$ and R_O. In this case the sum of Eq. (29) reduces to one term with the result that $\bar{E}_O^\Gamma X$ becomes a constant P_O almost everywhere, i.e., from Eqs. (29) and (30) one has

$$(\bar{E}_O^\Gamma X)(\theta) = \int_{R_O} X(\varphi)\, d\bar{P}_O(\varphi) = \langle X\rangle_{P_O} \tag{31}$$

holds P_O almost everywhere. Here and in the following $\langle X\rangle_P$ denotes the expectation value of X relative to P. From this, one has the following theorem.

Theorem 3 (Indecomposability)[15]*:* The following are equivalent:

(a) T' is P_O indecomposable and $\bar{P}_{O_n} \to \bar{P}_O$ on $A(R_O)$;

(b) $\lim_n \bar{X}_n$ exists and is a constant P_O almost everywhere for every nonnegative measurable X;

(c) for each $E_R \in A(R_O)$, $I_{E_R} = \bar{P}_O E_R$, P_O almost everywhere.

Again the proof will not be given here as it is given elsewhere.[15]

In what follows $\bar{X}(\theta) = \cdots$ is equivalent to the longer statement that $\lim_n \bar{X}_n(\theta)$ exists and equals $\cdots$.

Corollary 1: If T, P, and O satisfy the conditions of Theorem 2 and the ergodic hypothesis holds, then for every nonnegative measurable function X

$$\bar{X}(\theta) = \int_{R_O} X(\varphi)\, d\bar{P}_O(\varphi) = \langle X\rangle_{P_O},$$

P_O almost everywhere.

Proof: Immediate from Theorems 2 and 3 and Eqs. (27) and (31).

Corollary 2: If T, P, and O satisfy the conditions of Theorem 2, the ergodic hypothesis holds, and $T_{O'}$ is P_O measure preserving, i.e., for each $E_R \in A(R_O)$

$$P_O T_{O'} E_R = P_O E_R, \tag{32}$$

then, for every nonnegative measurable X,

$$\bar{X}(\theta) = \int_{R_O} X(\theta)\, dP_O(\theta) = \langle X\rangle_{P_O}.$$

Proof: Immediate from Corollary 1 and Eqs. (23) and (24).

Corollary 3: If P, T, and O satisfy the conditions of Theorem 2, then,for every nonnegative measurable X, the formula

$$\bar{X}O\varphi = (\bar{E}_O^\Gamma X)(O\varphi)$$

holds P_D almost everywhere on D_O. If also O is such that $PD_O = 1$, then this formula holds P almost everywhere on Ω. The same conclusions apply to the formulas $\bar{X}O\varphi = \langle X\rangle_{P_O}$ and $\bar{X}O\varphi = \langle X\rangle_{P_O}$ if the ergodic hypothesis holds and also $T_{O'}$ is P_O measure preserving, respectively.

Proof: From the ergodic theorem and Eqs. (27), (22), (21), (16), and (15), one has

$$\begin{aligned} 1 &= P_O[\varphi \mid \bar{X}(\varphi) = (\bar{E}_O^\Gamma X)(\varphi)] \\ &= P_D[\varphi \mid \bar{X}O\varphi = (\bar{E}_O^\Gamma X)(O\varphi)] \\ &= P[\varphi \mid \bar{X}O\varphi = (\bar{E}_O^\Gamma X)(O\varphi)], \end{aligned}$$

where the last equality holds if and only if $PD_0 = 1$. Application of the same argument to the formulas of Corollaries 1 and 2 gives the desired result.

B. Main Results

The following metatheorems include the conditions that X, O, and T are τ-definable. To see what this means, we first consider T. Entirely similar definitions hold for O and X.

To say that T is τ-definable means that there is a formula $A(\varphi, \theta)$ in $L(\tau)$ with the following properties: (a) the existence and uniqueness conditions,

$$\exists\varphi\exists\theta A(\varphi, \theta)$$

and $\forall\varphi\forall\theta\forall\psi\,(A(\varphi, \theta) \wedge A(\varphi, \psi) \Rightarrow \theta = \psi)$ are theorems of τ, and (b) the defining formula $\forall\varphi\forall\theta\,(T(\varphi) = \theta \Leftrightarrow A(\varphi, \theta))$ is true for T. [Other conditions in the definition of T, such as the equality between Ω and the domain and range sets of T and the $A(\Omega)$ measurability of T, are given by other formulas.] If T is τ-definable, then one can extend τ and $L(\tau)$ by adding a symbol T to $L(\tau)$ as the name of T and the defining formula with T replacing T as a defining axiom (Sec. II).

In a strict sense these formulas are not yet formulas of $L(\tau)$ as φ and θ are informal elements and are not symbols of $L(\tau)$. This is remedied by replacing $A(\varphi, \theta)$ by $A(x, y) \wedge \text{seq}\ (x) \wedge \text{seq}\ (y)$, $\forall\varphi B(\varphi \cdots)$ by

$$\forall x\ (\text{seq}\ (x) \Rightarrow B(x \cdots)),$$

and $(\exists\varphi B(\varphi \cdots)$ by $\exists x\ (\text{seq}\ (x) \wedge B(x \cdots))$ in the above. Here $B(\varphi \cdots)$ denotes any formula with at least φ free, and seq (x) denotes the defining relation in set theory for "x is an infinite sequence (set of ordered pairs) of natural numbers."

The definition of τ-definability for O is exactly the same as the difference between O and T (as a special case of O) appears in the other defining conditions.

The τ-definability of X is given by replacing θ and ψ in the above by real number variables. Corresponding replacements of seq $(-)$ by Re $(-)$ must also be made.

Metatheorem 4: If P, T, O, and X satisfy the conditions of the ergodic theorem and if T, O, and X are τ-definable, then the formula obtained by replacing X, T, and O by their respective names **X**, **T**, and **O** in $\bar{X}(O\varphi) = \bar{E}_O^{\Gamma} X(O\varphi)$ is a formula in (the extended) $L(\tau)$ with P and φ free. Similarly, the formulas $\bar{X}O\varphi = \langle X\rangle_{P_0}$ and $\bar{X}O\varphi = \langle X\rangle_{P_0}$ with X, T, and O τ-definable are formulas in $L(\tau)$ with P and φ free.

Proof: The formula $\bar{X}(O\varphi) = \bar{E}_O^{\Gamma} X(O\varphi)$ can be converted into a formula of τ, as it is tedious but straightforward to show that the various concepts involved of limit, minimal σ-field on the E_{nl}, $A(\Omega)$-measurable functions, integral, etc., are all definable in τ. Thus, by use of the process of extension by definitions Eqs. (1)–(10) and the replacement of X, T, and O by their respective names, one obtains a formula $q(P, \varphi)$ in the extended $L(\tau)$ with P and φ free. [As noted before, $q(P, \varphi)$ is strictly not in $L(\tau)$, but the corresponding $q(x, y) \wedge \Pr(x) \wedge \text{seq}(y)$ is in $L(\tau)$.]

By the same arguments one shows that $\bar{X}O\varphi = \langle X\rangle_{P_0}$ and $\bar{X}O\varphi = \langle X\rangle_{P_0}$ correspond to formulas in the extended $L(\tau)$. QED

From this, one obtains directly the desired result in the following metatheorem.

Metatheorem 5: Let P be a probability measure and ψ a sequence of natural numbers such that $E(\tau, P, \psi)$ [Eq. (12)] holds. Let T, O, and X satisfy the conditions of Theorem 2 and be τ-definable. Also let P and O be such that $PD_O = 1$. Then the following hold:

(1) If $\lim_{n=\infty} \bar{P}_{On} = \bar{P}_O$ exists, then

$$\bar{X}O\psi = \bar{E}_O^{\Gamma} X(O\psi); \tag{33}$$

(2) if, further, Γ is, up to a P_O equivalence, generated by a countable partition $\{C_i\}$ of R_O, then

$$\bar{X}O\psi = \sum_i (\bar{E}_{OC_i} X) I_{C_i}(O\psi), \tag{34}$$

with $\bar{E}_{OC_i} X$ given by Eq. (30) and $\bar{P}_O C_i > 0$ for that C_i for which $O\psi \in C_i$;

(3) if, further, the ergodic hypothesis holds, then

$$\bar{X}O\psi = \int_{R_O} X(\varphi)\, d\bar{P}_O(\varphi) = \langle X\rangle_{\bar{P}_O}; \tag{35}$$

(4) if, further, T is P_O measure preserving, then

$$\bar{X}O\psi = \int_{R_O} X(\varphi)\, dP_O(\varphi) = \langle X\rangle_{P_O}. \tag{36}$$

Proof: Since T, O, and X are τ-definable, one can, in Eqs. (33)–(36),replace T, O, and X by their respective names **T**, **O**, and **X** and add the defining relations as axioms to τ. By Metatheorem 4, each of the resultant formulas, with φ replacing ψ, is a formula $q_i(P, \varphi)$, $i = 1 \cdots 4$, in $L(\tau)$ with P and φ free. In Eq. (34) the $\sum_i$ is equivalent to a binding of the variable C_i.

Furthermore, by Theorems 2 and 3 and Corollaries 1–3, if the conditions given in (1)–(4) hold, then the appropriate formula $q_i(P, \varphi)$ is true P almost everywhere {i.e., $P[\varphi \mid q_i(P, \varphi)] = 1$}, and, by the definition of P (or the definition of Cauchy limit and the measurability of X, T, and O), $[\varphi \mid q_i(P, \varphi)] \in A(\Omega)$. Thus from the definition of $E(\tau, P, \psi)$, if the conditions given in (1)–(4) hold, then one has $q_i(P, \psi)$. [Note that in part (2) there exists a C_i such that $P_O C_i > O$ and $O\psi \in C_i$. For suppose that $O\psi \in C$ but that $\bar{P}_O C = O$. Then, by Eqs. (29) and (30), $\bar{X}O\psi$ would not exist, contrary to the statement of part (1).] QED

Some comments to clarify the meaning of this metatheorem are in order. First one notes that if the general hypotheses of the metatheorem are satisfied and $\lim_n \bar{P}_{On} = \bar{P}_O$ exists, then $\lim_n \bar{X}_n(O\psi)$ exists and equals $(\bar{E}_O^{\Gamma} X)(O\psi)$, with $(\bar{E}_O^{\Gamma} X)(\varphi)$ given by Eq. (28). Now this statement, Eq. (33), holds for all four parts of the metatheorem. Parts (2)–(4) are resultant statements which hold, in addition, if the general definition of $\bar{E}_O^{\Gamma} X$ as the Radon–Nikodym derivative of a set function can be made more explicit.[14,15]

In particular, if T, P, and O are such that, up to a P_O equivalence, Γ is generated by a countable partition of R_O, then, outside of P_O null sets in Γ, $(\bar{E}_O^{\Gamma} X)(\varphi)$ has at most countably many distinct values. [The values of $(\bar{E}_O^{\Gamma} X)(\varphi)$ are constant within each generator of Γ irrespective of how many there are.] Then part (2) of the metatheorem says that there exists a generator C such that $P_O C > O$ and $O\psi \in C$ and $\bar{X}O\psi = \bar{E}_{OC} X$ given by Eq. (30). The formal statement of this,

$$\exists C\,(P_O C > O \wedge O\psi \in C \wedge \bar{X}O\psi = \bar{E}_{OC} X), \tag{37}$$

with C a variable ranging over all generators of Γ, is a formula $q(P, \psi)$ of $L(\tau)$ with P and ψ free, which is equivalent to Eq. (34) and associated statements and must be true at ψ.

If, further, the ergodic hypothesis holds, then, outside P_O null sets in Γ, $(\bar{E}_O^{\Gamma} X)(\varphi)$ has only one value, $\langle X\rangle_{P_O}$, which is obtained from Eq. (30) by setting $C_i = R_O$ (recall that $T' R_O = R_O$). Thus in this case $\bar{X}$ has, P_O almost everywhere, only one value and thus must assume that value at $O\psi$. This is what part (3)

says. Finally, part (4) of the metatheorem is a consequence of the fact that if T is P_O measure preserving [Eq. (32)], then by Eqs. (23) and (24) $\bar{P}_O = P_O$ on $A(R_O)$.

Finally, there are some more general comments to be made. First one notes that the definition of $E(\tau, P, \psi)$ tells one, for all formulas q in $L(\tau)$, if . . . , then $q_P(P, \psi)$. The definition says nothing about how one is to determine for each q whether or not the "if" conditions are satisfied. In particular, one wants to know whether or not a formula $q_P(P, \varphi)$ is true P almost everywhere.

The axioms of τ play a role here, as it is possible to prove from them theorems which state that certain formulas satisfy the conditions in the definition of $E(\tau, P, \psi)$. As an example, for formulas of the general form of Eqs. (33)–(36), the ergodic and indecomposability theorems give conditions on O, P, and T which must be satisfied if the formulas are to be true P almost everywhere. However, in this case the problem then arises as to how one knows whether or not the conditions on O, P, and T are satisfied.

The requirement that X, O, and T be τ-definable is important. For suppose that O were not τ-definable. Then the formula $XO\psi = (\bar{E}_O^{\Gamma}X)(O\psi)$ would be a formula $q(O, P, \varphi)$ with O, P, and φ free and would not satisfy the defining conditions in the definition of $E(\tau, P, \psi)$. Thus there would be nothing preventing ψ from lying in the P null set of the above formula [i.e., nothing would prevent $\bar{X}O\psi \neq (\bar{E}_O^{\Gamma}X)(O\psi)$ from being true].

A more important reason for the τ-definability of T, O, and X is that sufficient conditions are thereby imposed on the class of allowed T, O, and X to satisfy the existence requirement (Metatheorem 1) for $E(\tau, P, \psi)$. To see this, suppose that the definition of $E(\tau, P, \psi)$ admitted all formulas of the type $\bar{X}O\psi = \langle X\rangle_{P_O}$ for *all* O, not just the τ-definable ones. Formally, this is achieved by the well-known method of addition of constants[6] to $L(\tau)$. That is, for each O, one adds a constant symbol $\mathbf{O}$ to $L(\tau)$ as a name for O without adding defining formulas (which do not exist for most O). Then for each O, and τ-definable T and X, if O, T, X, and P satisfy the conditions of the ergodic theorem and Corollaries 1 and 3, the formula $\bar{X}\mathbf{O}\varphi = \langle X\rangle_{P_O}$ satisfies the conditions in the definition of $E(\tau, P, \psi)$ now given for the greatly extended $L(\tau)$. Thus $\bar{X}\mathbf{O}\psi = \langle X\rangle_{P_O}$ must hold for *all* such O.

However, in this case one can see that, at least for some important cases, there exist no ψ which satisfy this requirement. [The similar case of the failure of the existence requirement for $E(\tau, P, \psi)$ under arbitrary additions of names of sequences to $L(\tau)$ has been noted elsewhere.][9] To see this, let ψ be any sequence, and let n be a number which ψ contains an infinite number of times. Let $X = I_{E_{n0}}$, the characteristic function for the event E_{n0} [Eq. (11)], and T be the one-sided shift operator [Eq. (19)], and P a nontrivial product measure such that $O < PE_{n0} < 1$. Define O to be that subsequence selection procedure O_g [Eq. (18)], such that the corresponding place label function $g: N \to N$ is such that $\psi(g(j)) = n$ for each j. It is easy to show[13,16] that $\bar{X}O_g\varphi = PE_{nO} < 1$, P almost everywhere, but that $\bar{X}O_g\psi = 1$ and thus $\bar{X}O_g\psi \neq \langle X\rangle_{P_0} = \langle X\rangle_{P_0}$.

Entirely similar arguments can be used in the case of the τ-definability of T and X. In essence the argument used here is an adaptation of the argument, used to show the nonexistence of random sequences when the definition was given in terms of the classical totality of subsequence selection procedures.[1,11–13]

V. AGREEMENT BETWEEN THEORY AND EXPERIMENT

A. Physical Theories

Here the same type of construction as was given in I is used. Let Q be a mapping $Q: N \to G \times K$, where N is the set of natural numbers, G is the set of instructions for all preparation acts, and K is the set of instructions for all observation acts. Each instruction g in G and k in K is considered to be a sentence of arbitrary but finite length, written in some suitable informal language. Also G and K contain *all* such instructions, not just those appropriate to physics. Also (I, Appendix) it is assumed that the result of carrying out any observation act k in K yields (a symbol for) a natural number and not a real number.

As before,[1] the triple (Qst) denotes an infinite string of preparation and observation acts where $s: N \to N$ and $t: N \to N$ are functions from N to N. For each j, $Q(j) = (g_j, k_j)$ is an ordered pair of instructions for a preparation and observation act, respectively. $s(j)$ and $t(j)$ give the space and time positions (relative to some space–time ordering procedure) at which $Q(j)$ is to be carried out. It is assumed that each k in K includes instructions for the space–time positioning relative to all preparation acts g in G. The triple $Q(j)$, $s(j)$, $t(j)$ thus describes the jth single measurement in the sequence, the carrying out of which will yield a natural number, and (Qst) describes an infinite sequence of single measurements, or an experiment.

It is to be emphasized that the restriction of all k's in K to include relative space–time positionings is made purely to conserve on notation and is entirely inessential. Clearly one can replace (Qst) by $(Qs_1t_1s_2t_2)$,

where s_1 and t_1 are functions which give the space–time positions for carrying out each preparation act in Q and s_2 and t_2 give the space–time positions for carrying out each observation act in Q. In this case such a restriction is not necessary as the more cumbersome notation $s_1t_1s_2t_2$ carries all the necessary information.

It is also to be noted that, in essence, s and t describe orderings of the elements of Q with respect to other background observations as seen by an observer. They do not assign a particular space–time metric to the elements of Q. Such an assignment will not be considered here as it is outside the scope of this paper. Suffice it to say that, in common with other authors,[17,18] it is felt to be desirable to clearly separate the order aspects of space–time from the metric aspects.

Let $[Qst]$ denote the set of all infinite strings of instructions which an observer can actually carry out. Clearly this is a subclass of the class of all possible triples, as some instructions in G or K may be incomplete, or impossible to carry out, or correspond to nonterminating procedures, etc. Also t must be nondecreasing as one cannot do the $(n+1)$th single measurement before the nth.

It is to be emphasized that the detailed nature of $[Qst]$ is irrelevant to this work. If desired, $[Qst]$ can be considered to be the class of all infinite sequences of single measurements, or some other sufficiently inclusive description of sequences of arbitrary empirical acts. It is not necessary for this work that the elements of $[Qst]$ be sequences of instructions.

The only properties of $[Qst]$ which are really necessary here are (1) $[Qst]$ must be sufficiently complete to include all infinite sequences of single measurements which are properly the domain of physics and (2) the elements of $[Qst]$ must correspond to *infinite* sequences of empirical acts which yield *infinite* sequences of (symbols for) natural numbers.

The requirement that the outcome of a single measurement be a natural number is used because, as was discussed in I, the outcome of a single measurement corresponds to a natural number, not a real number. As far as the probability theoretic aspects of this work are concerned, one can just as easily consider infinite sequences of real numbers. It is not clear, though, if one can carry over the definition of $E(\tau, P, \psi)$ and the resulting metatheorems to apply to infinite sequences of real numbers.

For each (Qst) in $[Qst]$, define ψ_{Qst} to be the sequence obtained by actually carrying out or doing (Qst). Finally, let U be a mapping with domain in $[Qst]$ and range in the set of all probability measures defined on $A(\Omega)$.

In this work a physical theory will be considered to correspond, in essence, to some mapping U. That is, a physical theory is to be thought of as some system such that, for each experiment (Qst) in its domain, the theory generates a probability measure $U(Qst)$ on $A(\Omega)$. {From now on, the triples (Qst) in $[Qst]$ are referred to as experiments.}

Now, in general, different physical theories correspond to different mappings U and thus will assign different probability measures to the same (Qst). Thus not all physical theories are valid, and there are clearly restrictions on the class of all mappings U in order that U be a valid physical theory. (Since each physical theory, as an extended system, is assumed to generate a unique mapping U, a physical theory will often be identified with the mapping U which it generates.)

B. Agreement between Theory and Experiment

One important restriction on a theory U is that if U is valid, it must agree with experiment. It is proposed that the statement of agreement between a physical theory U and experiment, relative to τ, $A(\tau, U)$ be defined by

$$A(\tau, U) \equiv \forall Qst\ [U(Qst) \text{ defined} \Rightarrow E(\tau, U(Qst), \psi_{Qst})], \quad (38)$$

with $E(\tau, U(Qst), \psi_{Qst})$ given by Eq. (12).

In words this definition says the following: U τ-agrees with experiment if and only if, for each experiment (Qst) in the domain of U, any property of sequences which is expressible by a formula in $L(\tau)$ with one free sequence variable and which may contain the measure $U(Qst)$ and which is true $U(Qst)$ almost everywhere is a property possessed by ψ_{Qst}.

An equivalent statement of this definition is that, for each experiment (Qst) in the domain of U, the class of tests for agreement of the theory U at (Qst) is the class of τ-definable properties of sequences [where $U(Qst)$ can occur in the defining relation] possessed by $U(Qst)$ almost all sequences. The definition then says that each property in the class must be possessed by the sequence obtained by actually carrying out (Qst), i.e., by ψ_{Qst}.

Note that the class of tests for agreement at any experiment clearly depends on the experiment. The reason is that, in general, a τ-definable property possessed by $U(Q's't')$, almost all sequences, need not be possessed by $U(Qst)$, almost all sequences.

Now, if Eq. (38) is to be an acceptable definition, it must be shown that the class of τ-definable properties of sequences is sufficiently broad to include the full intuitive meaning of agreement between theory and experiment. This was gone into in some detail in I.

There it was seen that, for the more restrictive H (approximately second-order arithmetic) replacing τ in $A(\tau, U)$, the proposed definition did include much of the intuitive meaning of agreement between theory and experiment.

However, it was also seen that, for H as defined in I, there were many important properties which are part of the intuitive meaning of agreement between theory and experiment and were not included in the definition. These included the limit properties of sequences which must be related to expectation values given by the physical theory.

One of the main purposes of this paper is to show that these properties excluded by the definition given in I are included in the stronger definition given here with $\tau =$ ZF set theory. In fact, one has the following metatheorem.

Metatheorem 6: Let U agree with experiment with respect to τ. Let (Qst) be any instruction sequence in the domain of U. Let T be any $A(\Omega)$-measurable transformation with domain and range equal to Ω. Let O be an measurable transformation with domain $D_O \subseteq \Omega$ and range $R_O \subseteq \Omega$ such that (a) $D_O \in A(\Omega)$ and $R_O \in A(\Omega)$, (b) $U(Qst)D_O = 1$, and (c) $T'R_O = R_O$. Let X be any nonnegative, measurable function from R_O to the real numbers.

Then, if T, O, and X are τ-definable, the following hold:

(1) If $\lim_n \overline{U(Qst)}_{On} = \overline{U(Qst)}_O$ exists, then

$$\bar{X}O\psi_{Qst} = (\bar{E}_O^{\Gamma}X)(O\psi_{Qst}), \tag{39}$$

with Γ the sub σ-field of T' invariant sets in $A(R_O)$ and $\bar{E}_O^{\Gamma}X$ defined by Eqs. (28)–(30) with $\overline{U(Qst)}_O$ replacing $\bar{P}_O$.

(2) If also $U(Qst)$, O, and T are such that, up to a $U(Qst)_O$ equivalence, Γ is generated by a countable partition $\{C_i\}$ of R_O, then

$$\bar{X}O\psi_{Qst} = \sum_i (\bar{E}_{OC_i}X)I_{C_i}(O\psi_{Qst}), \tag{40}$$

with $U(Qst)_O C_i > O$ for that C_i for which $O\psi_{Qst} \in C_i$ and $\bar{E}_{OC_i}X$ given by

$$\bar{E}_{OC_i}X = \frac{1}{\overline{U(Qst)}_O C_i}\int_{C_i} X(\varphi)\, d\overline{U(Qst)}_O(\varphi). \tag{41}$$

(3) If also $U(Qst)$ and O are such that the ergodic hypothesis holds, then

$$\bar{X}O\psi_{Qst} = \int_{R_O} X(\varphi)\, d\overline{U(Qst)}_O(\varphi) = \langle X\rangle_{\overline{U(Qst)}_O}. \tag{42}$$

(4) If also $U(Qst)$ and O are such that T is $U(Qst)_O$ measure preserving,

$$\bar{X}O\psi_{Qst} = \int_{R_O} X(\varphi)\, dU(Qst)_O(\varphi) = \langle X\rangle_{U(Qst)_O}. \tag{43}$$

Proof: By the definition of τ-agreement between theory and experiment [Eq. (38)], $E(\tau, U(Qst), \psi_{Qst})$ holds for each Qst in the domain of U. By hypothesis, T, O, X, and $U(Qst)$ satisfy the conditions of Metatheorem 5. Since parts (1)–(4) are merely restatements of the corresponding parts of Metatheorem 5 for the measure $U(Qst)$ and the sequence ψ_{Qst}, one has immediately that (1)–(4) hold. QED

In order to better understand this metatheorem and the definition of τ-agreement between theory and experiment, some examples are given. Let O be the identity operator, T the one-sided shift operator [Eq. (19)], and $X = \sum_n nI_{E_{n0}}$, with $I_{E_{n0}}$ the characteristic function for the set E_{n0} [Eq. (11)]. This particular X and T occur so often that the resulting $\bar{X}$ is denoted here by $\bar{M}$, where $\bar{M}\varphi$ is the limit mean of φ defined by

$$\lim_{n\to\infty} \frac{1}{n}\sum_{j=0}^{j=n-1} \varphi(j).$$

Now let the physical theory U and the experiment (Qst) be such that, among other things, $\overline{U(Qst)}$ exists for the one-sided shift, the ergodic hypothesis is not valid, and the Γ is almost surely generated by a countable partition of Ω ($R_O = \Omega$). This means that U predicts that the limit mean of the sequence obtained by actually doing (Qst) will be equal to some real number in a set of at most countably many real numbers. Furthermore, the prediction is that the particular value assumed by the limit mean is determined by the requirement that $\psi_{Qst} \in C$, Eq. (37) for some non-$U(Qst)$ null generator C of Γ. Also, each number in the set is equal to $\sum_n n\overline{U(Qst)}(E_{n0} \mid C)$, where $\overline{U(Qst)}(E_{n0} \mid C)$ is the conditional probability of the event E_{n0}, given that C has occurred [Eq. (41)], where C is a non-$U(Qst)$ null generator of Γ. Note that U does not predict what the value of $\bar{M}\psi_{Qst}$ will be. It states that $\bar{M}\psi_{Qst}$ must be in a particular set of real numbers and gives a condition for determining which number in the set equals $\bar{M}\psi_{Qst}$.

The metatheorem now says that if U τ-agrees with experiment, one must find, on actually doing (Qst), that this prediction is true. That is, one must find out that

$$\bar{M}\psi_{Qst} = \sum_n n\frac{\overline{U(Qst)}\,(E_{n0} \cap C)}{\overline{U(Qst)}C} \tag{44}$$

is true for that non-$U(Qst)$ null generator C for which $\psi_{Qst} \in C$.

If, however, the ergodic hypothesis is valid for T and $U(Qst)$, then this means that U predicts that $\bar{M}\psi_{Qst}$ lies in a set containing only one real number, $\sum_n n\overline{U(Qst)}E_{n0}$, and that $\psi_{Qst} \in C^\circ$ where, up to a $U(Qst)$ equivalence, $C^\circ = \Omega$ and Γ is generated by C° and $\varnothing$. Clearly, here U does predict what the value of $\bar{M}\psi_{Qst}$ will be. By the metatheorem, if U τ-agrees with experiment one must find out, on doing Qst, that

$$\bar{M}\psi_{Qst} = \sum_n n\overline{U(Qst)}E_{n0} \tag{45}$$

is true [Eq. (42)].

Finally, if U says that (Qst) is such that $U(Qst)$ is T invariant, then, if U τ-agrees with experiment, the metatheorem says that one must find out that [Eq. (43)]

$$\bar{M}\psi_{Qst} = \sum_n nU(Qst)E_{nO} \tag{46}$$

is true.

{It should be recalled that, for any O, T, and X, not just those in this example, $\overline{U(Qst)}_O$ (if it exists) is a T-invariant measure [Eq. (32)] and $\bar{X}$, as a Γ-measurable function, is also T invariant. Thus, in this example one has $\overline{U(Qst)}E_{ni} = \overline{U(Qst)}E_{n0}$ and $\bar{M}T^j\psi_{Qst} = \bar{M}\psi_{Qst}$ independent of j.}

As another example let $O\varphi = F_m\varphi$, where, for all j and φ,

$$\begin{aligned}(F_m\varphi)(j) &= 1 \quad \text{if} \quad \varphi(j) = m \\ &= 0 \quad \text{if} \quad \varphi(j) \neq m,\end{aligned} \tag{47}$$

and T and X be as above. Clearly this O satisfies the conditions given in Metatheorem 6 as $D_O = \Omega$, $R_O = \Omega_{O,1}$, the set of all infinite O–1 sequences, $T'R_O = R_O$, and O is measurable. If U τ-agrees with experiment, all infinite sequences (Qst) of single measurements, for which T is ergodic with respect to $U(Qst)$, must be such that one finds, on carrying out (Qst), that

$$\bar{M}F_m\psi_{Qst} = \overline{U(Qst)}E_{m0}.$$

Here the facts that T commutes with F_m and that $U(Qst)F'_mE_{1j} = U(Qst)E_{mj}$ for each j have been used. If $U(Qst)$ and T satisfy an ergodic theorem but not the ergodic hypothesis, or if T is ergodic and measure preserving for $U(Qst)$, then relations similar to Eqs. (44) and (46) will be valid. [Note the relation $\bar{M}\varphi = \sum_m m\bar{M}F_m\varphi$ holds $U(Qst)$ almost everywhere as $\bar{M}F_m\varphi$ is the limit relative frequency for finding m in φ.]

As a somewhat more complex example, let $O\varphi = F_m\varphi \times F_nT^j\varphi$ and T and X be as before. Then if U τ-agrees with experiment and T is ergodic for $U(Qst)$, Metatheorem 6 gives

$$\begin{aligned}\bar{M}O\psi_{Qst} &= \bar{M}(F_m\psi_{Qst} \times F_nT^j\psi_{Qst}) \\ &= \overline{U(Qst)}\,(E_{m0} \cap E_{nj}).\end{aligned}$$

Here, too, T commutes with O and $O'E_{1k} = E_{mk} \cap E_{n,k+j}$ for each k. This O is an example of those O's through which correlations among the elements of (Qst) are described. Thus, if U and (Qst) are such that the formula $q(\varphi)$ of $L(\tau)$ defined by

$$q(\varphi) \equiv \bar{M}(F_m\varphi \times F_nT^j\varphi) = \bar{M}F_m\varphi \times \bar{M}F_n\varphi$$

is true $U(Qst)$ almost everywhere, then, if U τ-agrees with experiment, Metatheorem 6 requires that

$$\bar{M}(F_m\psi_{Qst} \times F_nT^j\psi_{Qst}) = \bar{M}F_m\psi_{Qst} \times \bar{M}F_n\psi_{Qst}$$

hold.

As a final example in which O does not commute with T, let T and X be as above and O be the subsequence selection procedure "select, in the natural order, all elements which follow the occurrence of a 1 by one place." Clearly O satisfies the conditions of Metatheorem 6. $D_O = [\varphi \mid \varphi$ contains an infinite number of 1's$] \subset \Omega$, $R_O = \Omega$, $T'R_O = R_O$, and one can easily show that $D_O \in A(\Omega)$ and O is measurable. Also O is clearly $L(\tau)$ definable; in fact, for each φ and j, $(O\varphi)(j)$ is φ-effectively calculable.[5,6]

If (Qst) is such that T is ergodic for $U(Qst)$ and also $U(Qst)D_O = 1$, then, if U τ-agrees with experiment, Metatheorem 6 gives that

$$\bar{M}O\psi_{Qst} = \sum_n n\overline{U(Qst)}_OE_{n0}$$

must be true. Note that since O does not commute with T, one cannot write $\overline{U(Qst)}_OE_{n0} = \overline{U(Qst)}O'E_{n0}$.

As an aid to further understanding the definition of τ-agreement between theory and experiment, consider the following question which often arises. Suppose one carries out an infinite sequence (Qst) of single measurements for which his theory predicts that the probability is one that each single measurement outcome in the sequence ψ_{Qst} is the same. The question then arises as to whether this means that, as a necessary condition of agreement of the theory at the experiment (Qst), one must find that ψ_{Qst} is a constant sequence or whether it is sufficient to require only that the limit relative frequency of finding the same outcome be equal to 1.

To see what the definition given here says about this question, suppose that the theory U and the experiment (Qst) are such that (a) $U(Qst)$ $[\varphi \mid \forall j(\varphi(j) = 1 \vee \varphi(j) = 0)] = 1$ and (b) the one-sided shift operator T [Eq. (19)], an Ω, is ergodic with respect to $U(Qst)$. Finally, (c) suppose one interprets the prediction that the probability is one that each outcome is the same to mean that

$$\overline{U(Qst)}E_{n0} = \delta_{1,n}, \tag{48}$$

that is, that the *limit ensemble* probability[19] for finding the outcome n for each single measurement equals 1 if $n = 1$ and equals 0 if $n \neq 1$. [Note that $\overline{U(Qst)}E_{nj} = \overline{U(Qst)}E_{n0}$ independent of j.]

From this one concludes the following: By Eqs. (12) and (38) and Metatheorem 6, if U τ-agrees with experiment, then from condition (a) one must have that $\forall j(\psi_{Qst}(j) = 1 \vee \psi_{Qst}(j) = 0)$ is true. That is, ψ_{Qst} must be a sequence of 0's and 1's only. Condition (b) with the ergodic and indecomposability theorems means that $\bar{M}\varphi = \sum_n n\overline{U(Qst)}E_{n0}$ is true on $\Omega, U(Qst)$ almost everywhere. Thus one must find out that

$$\bar{M}\psi_{Qst} = \sum_n n\overline{U(Qst)}E_{n0}. \tag{49}$$

Finally from condition (c), Eq. (48), one must find out that [Eq. (47)]

$$\bar{M}\psi_{Qst} = \bar{M}F_n\psi_{Qst} = 1 \tag{50}$$

or that the limit relative frequency for finding 1 in ψ_{Qst} is equal to 1.

For the case of individual outcomes, one notes that, from Eqs. (23) and (24), Eq. (48) is equivalent to [with $(T')^jE_{n0} = E_{nj}$]

$$\lim_{m\to\infty} \frac{1}{m} \sum_{j=0}^{m-1} U(Qst)E_{nj} = \delta_{1,n}. \tag{51}$$

Thus condition (c) says that the limit relative frequency that $U(Qst)E_{1j} = 1$, for $j = 0, 1, \cdots$, equals 1. It does not say that $U(Qst)E_{1j} = 1$ for all j. [For example, Eq. (51) is satisfied if $U(Qst)E_{1j} = 1$ unless $j = 2^l$ for some integer l. For these j's $U(Qst)E_{1j} < 1$.]

Thus, if U τ-agrees with experiment, the definition (38) requires for this case that, for each j for which $U(Qst)E_{1j} = 1$, one must find that $\psi_{Qst}(j) = 1$. For those j for which $U(Qst)E_{1j} < 1$, no prediction is made about the value of $\psi_{Qst}(j)$ [unless $U(Qst)E_{1j} = 0$, in which case (Metatheorem 2) one must find that $\psi_{Qst}(j) \neq 1$].

However, if one interprets the statement that a theory predicts that, with probability one, each outcome of a single measurement in (Qst) is the same to mean that T is also $U(Qst)$ measure preserving [i.e., that $U(Qst)E_{nj}$ is independent of j for all n], then from Eq. (51) one must have that $U(Qst)E_{1j} = 1$ for *all* j. Thus, in this case, if U τ-agrees with experiment, one must find that $\psi_{Qst}(j) = 1$ for all j.

Thus one sees that the definition of τ-agreement of a theory U with experiment does not give an unequivocal answer to the question raised. Rather it says that if U and (Qst) are such that T is ergodic and $\overline{U(Qst)}E_{mj} = \delta_{m,n}$ independent of j, the limit relative frequency for finding m in ψ_{Qst} equals 1 or that $\bar{M}F_m\psi_{Qst} = \delta_{m,n}$ [Eq. (47)]. The definition does not require that $\psi_{Qst}(j) = m$ for each j. However, if U and (Qst) are also such that T is $U(Qst)$ measure preserving, then if U τ-agrees with experiment one must find out that ψ_{Qst} is a constant sequence of m's.

One should note, however, that, even in the more general case with T ergodic but not measure preserving and $\overline{U(Qst)}E_{nj} = \delta_{n,m}$, not only does one require that for almost all j, $\psi_{Qst}(j) = m$, but the theory U provides a specific (and possibly incomplete) list of j values for which one must find out that $\psi_{Qst}(j) = m$. This is a consequence of the general property of the definition of τ-agreement between theory and experiment that any τ-definable subset of Ω which is a set of $U(Qst)$ measure 1 must include ψ_{Qst} as an element. This holds irrespective of whether the defining relation for the subset refers to a limit property or a nonlimit property of sequences.

With respect to randomness the following metatheorem, given in I, holds for the stronger definitions given here.

Metatheorem 7: Let U τ-agree with experiment. Then for all (Qst) in the domain of U, if $U(Qst)$ is a nontrivial product measure, ψ_{Qst} is τ-random.

Proof: Immediate from Eqs. (14) and (38).

C. Empirical Determinability of $U(Qst)$

It was pointed out in I that, on intuitive grounds, a physical theory which agrees with experiment should also have the property that at least for some (Qst) the probability $U(Qst)E$ for some events E in $A(\Omega)$ should be determinable from ψ_{Qst}. For example, suppose one computes the probability of observing spin up in the first single Stern-Gerlach measurement in an infinite sequence of such measurements of proton spin projections to be equal to $\frac{2}{3}$. (Each single measurement is made on a different proton.) One feels that under certain conditions one should be able to determine this probability by some limit mean operation on the infinite sequence of spin projection outcomes. In order to define this requirement and give sufficient conditions for it to be true, some definitions are necessary. The treatment is an extension and adaptation of that given in I to the stronger τ.

Let $A^\tau(\Omega)$ denote the subset of $A(\Omega)$ which contains all and only those subsets of Ω which are τ-definable; $A^\tau(\Omega)$ is a small subset of $A(\Omega)$ since $A^\tau(\Omega)$ has countably many elements, whereas $A(\Omega)$ has uncountably many. $A^\tau(\Omega)$ is a field of sets which clearly includes as a subset the minimal field $A^\tau_m(\Omega)$ containing the E_{nl}, with $n, l = 0, 1, \cdots$, as generators.

For each $F \in A^r(\Omega)$ let $O_F: \Omega \to \Omega_{0,1}$ be the transformation defined by

$$(O_F\varphi)(j) = 1 \quad \text{if} \quad T^j\varphi \in F$$
$$= 0 \quad \text{if} \quad T^j\varphi \notin F, \qquad (52)$$

for each φ and j, where T is the usual one-sided shift operator [Eq. (21)]. $\Omega_{0,1}$ is the set of all infinite sequences of 0's and 1's.

The statement that $U(Qst)$ is empirically determinable on $A^r(\Omega)$ from (Qst) is defined by

$$Dt(\tau, U, Qst)$$
$$\equiv \forall F\,[F \in A^r(\Omega) \Rightarrow \bar{M}O_F\psi_{Qst} \text{ exists and}$$
$$\bar{M}O_F\psi_{Qst} = U(Qst)F]. \qquad (53)$$

In words this definition says that for all τ-definable F, in $A(\Omega)$, the limit relative frequency that ψ_{Qst}, $T\psi_{Qst}$, $T^2\psi_{Qst}, \cdots$ is found in F exists and equals the probability of event F for the experiment (Qst) as given by the theory U.

Metatheorem 8 gives sufficient conditions for $Dt(\tau, U, Qst)$ to hold.

Metatheorem 8: Let U τ-agree with experiment. Then, for each (Qst) in the domain of U, sufficient conditions for $U(Qst)$ to be empirically determinable on $A^r(\Omega)$ from (Qst) are that the one-sided shift operator T be ergodic and measure preserving with respect to $U(Qst)$.

Proof: Assume that T is $U(Qst)$ measure preserving and ergodic on Ω. By Eq. (52) $D_{O_F} = \Omega$, $R_{O_F} = \Omega_{0,1}$, $T'R_{O_F} = R_{O_F}$, $U(Qst)D_{O_F} = 1$, and O_F is $A(\Omega)$ measurable and τ-definable if F is. Define X on $\Omega_{0,1}$ by $X = I_{E_{10}}$, the characteristic function for E_{10} [Eq. (II)]. Clearly X and T are τ-definable.

We must now show that for each τ-definable F:

(a) $\langle X\rangle_{\overline{U(Qst)_{O_F}}} = \overline{U(Qst)}F$ if the limit exists;

(b) if T is $U(Qst)$ measure preserving and ergodic on Ω, then T is $U(Qst)_{O_F}$ measure preserving and ergodic on $\Omega_{0,1}$.

Suppose we have shown the above propositions. By hypothesis, O_F, T, and X satisfy the general conditions of Metatheorem 6. Since T is $U(Qst)$ measure preserving and ergodic, $\overline{U(Qst)}F$ exists and $\overline{U(Qst)}F = U(Qst)F$. By hypothesis and (b) the conditions of part (4) of Metatheorem 6 are satisfied, and, using (a), one gets (with $\bar{M} = \bar{X}$ on $\Omega_{0,1}$)

$$\bar{M}O_F\psi_{Qst} = U(Qst)F.$$

Since this holds for all τ-definable F and obviously implies $\bar{M}O_F\psi_{Qst}$ exists, $Dt(\tau, U, Qst)$ is satisfied, and thus $U(Qst)$ is empirically determinable on $A^r(\Omega)$ from (Qst).

To prove (a), one notes that from Eqs. (23), (24), and the definition of X

$$\langle X\rangle_{\overline{(U\,Qst)_{O_F}}} = \lim_{n\to\infty} \frac{1}{n} \sum_{j=0}^{n-1} U(Qst)_{O_F}(T')^jE_{10}$$

if the limit exists.

Since $(T')^jE_{10} = E_{1j}$, and by Eqs. (52) and (15) $O'_FE_{nj} = (T')^jF$ if $n = 1$ and $O'_FE_{nj} = \Omega - (T')^jF$ if $n = 0$, one has that

$$U(Qst)_{O_F}(T')^jE_{10} = U(Qst)O'_FE_{1j} = U(Qst)(T')^jF.$$

Thus $\langle X\rangle = \overline{U(Qst)}F$ if the limit exists.

To prove (b), we first show that O_F and T commute. From Eq. (52) one has that $(O_FT\varphi)(j) = 1$[0] if $T^j \cdot T\varphi \in F$ [$\notin F$]. Since $T^j \cdot T\varphi = T^{j+1}\varphi$, one has, from Eqs. (52) and (19), $(O_FT\varphi)(j) = (O_F\varphi)(j+1) = (TO_F\varphi)(j)$ for all j and φ in Ω.

Thus, for any $E \in A(\Omega_{0,1})$, one has

$$U(Qst)_{O_F}T'E = U(Qst)O'_FT'E = U(Qst)T'O'_FE$$
$$= U(Qst)O'_FE = U(Qst)_{O_F}E,$$

and thus T is $U(Qst)_{O_F}$ measure preserving.

To prove the $U(Qst)_{O_F}$ ergodicity of T, one must show that $\overline{U(Qst)}_{O_F}$ exists and that T is $U(Qst)_{O_F}$ indecomposable. The former follows from the commutativity of O_F and T for every $E \in A(\Omega_{0,1})$:

$$\lim_{n\to\infty} \frac{1}{n} \sum_{j=0}^{n-1} U(Qst)O'_F(T')^jE$$
$$= \lim_{n\to\infty} \frac{1}{n} \sum_{j=0}^{n-1} U(Qst)(T')^jO'_FE,$$

which exists by hypothesis.

The latter follows from the $U(Qst)$ indecomposability of T. For suppose there were another T-invariant set $C \in A(\Omega_{0,1})$ besides $\Omega_{0,1}$ itself and $\varnothing$ (the empty set) with $O < U(Qst)_{O_F}C = U(Qst)O'_FC < 1$. Then, using the commutativity of O_F and T, we see that $T'O'_FC = O'_FT'C = O'_FC$ or O'_FC is a T-invariant set in $A(\Omega)$ which is $U(Qst)$ inequivalent to Ω or $\varnothing$. But this contradicts the $U(Qst)$ indecomposability of T. QED

This metatheorem gives sufficient conditions which must be satisfied for one to determine the measure on $A^r(\Omega)$ by computing $\bar{M}O_F\psi_{Qst}$ for each $F \in A^r(\Omega)$. The question arises whether the measure so determined can be extended to the full σ-field $A(\Omega)$. By an extension theorem[14,15] such an extension is possible, as $A^r(\Omega)$ is a field, and the extension is unique. In fact, such an extension is possible from the smaller minimal field, $A^r_m(\Omega)$ over the E_{nl} for $n, l = 0, 1, \cdots$. In this

case not only are elements of $A^{\tau}_{m}(\Omega)$ τ-definable, as in $A^{\tau}(\Omega)$, but it is ψ_{Qst}-effectively decidable[5] whether or not ψ_{Qst}, $T\psi_{Qst}$, $\cdots$, etc., belong to each set in $A^{\tau}_{m}(\Omega)$.

The above also suggests a partial answer to the following problem: Given any outcome sequence ψ_{Qst}, (a) can one generate or determine a probability measure from it and (b) does $E(\tau, P, \psi_{Qst})$ hold for the measure P so determined? Clearly sufficient conditions (using the extension theorem) for an affirmative answer to part (a) are that $\bar{M}O_{F}\psi_{Qst}$, considered as a function of F, satisfies the axioms of a probability measure. That is, for each $F \in A^{\tau}_{m}(\Omega)$, (1) $\bar{M}O_{F}\psi_{Qst}$ exists, (2) $0 \leq \bar{M}O_{F}\psi_{Qst} \leq 1$, (3) $\bar{M}O_{\Omega-F}\psi_{Qst} = 1 - \bar{M}O_{F}\psi_{Qst}$, (4) if F_i with $i = 1, 2, \cdots, n$ are all in $A^{\tau}_{m}(\Omega)$ and are pairwise disjoint, then

$$\bar{M}_{O\cup_{i\leq n}F_i}\psi_{Qst} = \sum_{i}^{n} \bar{M}O_{F_i}\psi_{Qst},$$

and (5) we have countable additivity, i.e., if the F_i with $i = 1, 2, \cdots$ are pairwise disjoint, then

$$\bar{M}O_{\cup_i F_i}\psi_{Qst} = \textstyle\sum_i \bar{M}O_{F_i}\psi_{Qst}.$$

Note that in this case $\bigcup_i F_i \notin A^{\tau}(\Omega)$ is possible even if all the F_i are in $A^{\tau}_{m}(\Omega)$.

Part (b) is a much deeper question whose answer would be the subject of another paper. It appears to be related to the consistency of such a determination as given in the answer to part (a).

VI. DISCUSSION

Metatheorem 6 and the associated discussion show that, with τ = ZF set theory, the definition of τ-agreement of theory with experiment is quite powerful. It includes relations between many limit properties of outcome sequences and expectations given by the physical theory as well as many other relations. There do not appear to be any properties which, on intuitive grounds, such a definition should have and which are not included in Eqs. (12) and (38).

In spite of this, it is an open question whether or not the definition as given here will ultimately have to be changed. For example, one might feel that, from the discussion following Metatheorem 6, the definition should be limited to including limit properties only of sequences. It is possible that one should never require, as a test for agreement between theory and experiment, that $\psi_{Qst}(j) = m$ even if $U(Qst)E_{mj} = 1$. In this case the definition given here would have to be altered to include a precise definition of a "limit" property.

Another possibility is that it might be necessary to alter or extend τ in some fashion. For example, one might want τ to include a formal language and formal theory for the (Qst) as well, or one might want to include properties of sequences which are expressed by formulas which include quantifiers over τ-definable properties of sequences. At present, little can be said about these alternatives other than that the definition given here seems sufficient.

Furthermore, Metatheorems 7 and 8 show that the definition given here has some intuitively satisfying properties. Intuitively one expects that any outcome sequence obtained from an experiment described by a product measure is random. This is just what Metatheorem 7 says. Also for such measures one feels that the measure should be determinable from the outcome sequence. This is included in Metatheorem 8, since a product measure is ergodic and invariant with respect to the shift operator.

One also notes that the requirement that U τ-agree with experiment is essential to Metatheorems 6–8. The reason is that without this requirement there is no reason why the conclusions of the metatheorems should hold for ψ_{Qst} even if $U(Qst)$ has the properties required in the hypotheses. The necessary connection is provided by Eqs. (12) and (38), whose validity is a consequence of U τ-agreeing with experiment.

As was noted in I, problems arise when one attempts to apply the results of this work to quantum mechanics. The main problem seems to be whether or not quantum mechanics is a theory of the type discussed here.[1] That is, the problem is whether or not, for each infinite sequence Qst of single measurements, quantum mechanics assigns a measure $U(Qst)$ to the event algebra $A(\Omega)$. As was noted in I, some necessary conditions for quantum mechanics to make such an assignment are that states which contain an infinite number of systems be allowed (unless one can reuse systems in subsequent single measurements) and that the projection axiom[20] or some appropriate generalization be valid for all measurements. Whether or not quantum mechanics can be so extended is, at present, open, although some attempts at such extensions have been made.[21]

It is also to be noted that whether or not a theory is of the type U as discussed here is quite separate from the problem of how one tells for any experiment (Qst) whether or not the single measurements are "the same" and are "mutually independent." To see this, one notes that Metatheorem 6 does not tell one directly to check for agreement between theory and experiment by comparing limit empirical means with computed expectation values. Rather, it gives ergodic and measure preserving conditions which the measure $U(Qst)$

must satisfy with respect to a τ-definable T if such a comparison is a test, at the experiment (Qst), of agreement with theory. It does not tell one how to determine whether or not T and $U(Qst)$ actually satisfy these conditions.

Similarly Metatheorem 8 gives the ergodic and measure preserving conditions as sufficiency conditions for the empirical determinability of $U(Qst)$ from (Qst). It also does not tell one how to determine whether or not T and $U(Qst)$ actually satisfy these conditions. As is well known,[15,22] the general problem of determining when a transformation T is ergodic and invariant with respect to a measure is quite difficult.

A closely related aspect is that Metatheorems 6 and 8 give the ergodic and measure preserving conditions as sufficiency conditions. The question arises whether or not these conditions are also necessary. A proof of this appears to be quite difficult. One problem is that, in the definition [Eq. (12)] of $E(\tau, P, \psi)$, the implication is one sided rather than being an equivalence. Another difficulty is that the equivalence statements of the ergodic theorem (Theorem 2) and the decomposability theorem (Theorem 3) refer to all measurable functions on all sets in a σ-field, not just the τ-definable ones.

Finally, there are some other open questions which should be mentioned. First, there is the problem of whether theories U with nonempty domains exist. Although intuitively it seems that such theories exist, one would like to have a proof of this.

Second, the relationship between this work and physics in general, and quantum mechanics in particular, needs more work. Here, a general class of theories was considered, and some properties were defined and discussed as necessary conditions for any theory in the class to be valid. It was not necessary to consider the details of any theory other than the properties of the measure which it generated for each experiment. Thus, in this respect, this work is applicable to any empirical theory of the type discussed, not just physics.

However, suppose one takes the epistemological viewpoint that physical reality, in essence, is defined by the basic aspects of the knowledge acquisition process and that a goal of this process is the construction of a comprehensive physical theory which, among other things, must agree with experiment. In this case it is clear that the subject matter of this paper is closely related to physics at a very basic level indeed.

* Work performed under the auspices of the U.S. Atomic Energy Commission.

[1] P. A. Benioff, J. Math. Phys. **11**, 2553 (1970). Hereafter referred to as I.

[2] E. P. Wigner, Am. J. Phys. **31**, 6 (1963); "Remarks on the Mind Body Question" in *The Scientist Speculates*, I. Good and W. Heinemann, Eds. (Putnam, London, 1962).

[3] M. M. Yanase, Am. J. Phys. **32**, 208 (1964).

[4] A. Shimony, Am. J. Phys. **31**, 755 (1963); B. D'Espagnat, *Things, Structures and Phenomena in Quantum Physics*, Third International Congress for Logic Methodology and Philosophy of Science, Amsterdam, 1967, B. Van Rootseluar and J. F. Staal,Eds. (North-Holland, Amsterdam, 1968), pp. 377–84; P. A. Moldauer, Am. J. Phys. **32**, 172 (1964).

[5] H. Rogers, Jr., *Theory of Recursive Functions and Effective Computability* (McGraw-Hill, New York, 1967).

[6] J. Shoenfield, *Mathematical Logic* (Addison-Wesley, Reading, Mass., 1967).

[7] K. Kuratowski and A. Mostowski, *Set Theory* (North-Holland, Amsterdam, 1968).

[8] P. Suppes, *Axiomatic Set Theory* (Van Nostrand, Princeton, N.J., 1960).

[9] P. A. Benioff, "A Theorem on Undefinability," unpublished paper.

[10] A. H. Kruse, Z. Math. Logik Grundlagen Math. **13**, 299 (1967).

[11] P. Martin Löf, "On the Notion of Randomness," *Proceedings of Symposium on Proof Theory and Intuitionism*, Summer Institute, State University at Buffalo, New York, 1968, J. Myhill, A. Kino, and R. Vesley, Eds. (North-Holland, Amsterdam, 1970); Inform. Control **9**, 603 (1966).

[12] A. Wald, Ergeb. Math. Kolloquiums **8**, 38 (1937); *Selected Papers in Statistics and Probability* (McGraw-Hill, 1955), pp. 25–45; R. Von Mises, *Mathematical Theory of Probability and Statistics*, H. Geiringer, Ed. (Academic, New York, 1964), Chap. I.

[13] A. Church, Bull. Am. Math. Soc. **46**, 130 (1940); P. A. Benioff, "On Random Sequences," unpublished paper; D. W. Loveland, Trans. Am. Math. Soc. **125**, 497 (1966); Z. Math. Logik Grundlagen Math. **12**, 279 (1966).

[14] P. Halmos, *Measure Theory* (Van Nostrand, Princeton, N.J., 1950).

[15] M. Loève, *Probability Theory* (Van Nostrand, Princeton, N.J., 1963), 3rd ed.

[16] J. Doob, Ann. Math. **37**, 363 (1936).

[17] C. H. Randall, thesis, Rensselaer Polytechnic Institute, KAPL 3147, June 1966.

[18] D. Finkelstein, Phys. Rev. **184**, 1261 (1969).

[19] P. A. Benioff, J. Math. Phys. **9**, 514 (1968).

[20] J. Von Neumann, *Mathematical Foundations of Quantum Mechanics*, R. T. Beyer, Transl. (Princeton U.P., Princeton, N.J., 1955).

[21] H. Everett III, Rev. Mod. Phys. **29**, 454 (1957); J. A. Wheeler, *ibid.*, 463 (1957); B. S. DeWitt, in *Batelle Rencontres*, C. M. DeWitt and J. A. Wheeler, Eds. (Benjamin, New York, 1968), pp. 318–32.

[22] P. Halmos, *Ergodic Theory* (Chelsea, New York, 1956).

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 12, NUMBER 3 MARCH 1971

General Involutional Transformations and the Representation of $GL(n)$

T. S. SANTHANAM*
Physikalisches Institut der Universität Bonn, Bonn, Germany
AND
P. S. CHANDRASEKARAN AND NALINI B. MENON
Matscience, The Institute of Mathematical Sciences, Madras, India

(Received 15 July 1970)

The properties of general involutional matrices satisfying the equation $A^m = kI$, $k = \text{const}$, are studied. Generating equations for the induced representations of an arbitrary $(n \times n)$ matrix are explicitly written down. The special case when $n = 3$ is discussed in detail. It is shown that the conditions which make an arbitrary $(n \times n)$ matrix involutional leave any of its induced representation involutional. It is further shown that an involutional matrix in its self-representation can be expanded in the basis of generalized Clifford algebra with coefficients which are the generalized hyperbolic functions. Eigenvalues of induced matrices, in particular when they are involutional, are calculated.

1. INTRODUCTION

General involutional transformations which include homographic projective transformations (apart from sign) have wide applications in physics.[1] These are matrices satisfying the relation $A^m = kI, k = \text{const}$, of which a particular case is the set of Pauli matrices.

The case when the set of matrices A obey the generalized Clifford algebra C_n^m (GCA) defined by

$$e_i e_j = \omega e_j e_i, \quad i < j, i, j = 1, \cdots, n, \tag{1.1}$$

$$e_i^m = 1, \tag{1.2}$$

where ω is a primitive mth root of unity, has been studied exhaustively. The general mathematical formulation has been made by Morinaga and Nono,[2] Yamazaki,[3] and Morris,[4] while its relation to physics through the study of their specific representations has been made systematically by Ramakrishnan[5] and collaborators.[1] The present investigation, however, is on involutional matrices which satisfy Eq. (1.2) and may or may not satisfy Eq. (1.1). In this sense, Eq. (1.2) alone envelops a wider class of matrices than those implied by both the Eqs. (1.1) and (1.2). The case when $m = 2$ has been studied in detail by Kim[6] recently. The object of the present work is the study of general involutional matrices.

When $m = 2$, an involutional matrix has the general form

$$A^{(2)} = \begin{bmatrix} a & b \\ c & -a \end{bmatrix}, \tag{1.3}$$

except for trivial constant matrices where a, b, and c are arbitrary parameters. If this is regarded as an element of the general linear group in two dimensions, the matrix representation of $A^{(2)}$ as a transformation on a basis set of homogeneous polynomials of qth degree in two variables will yield a $(q + 1) \times (q + 1)$ involutional matrix with three arbitrary parameters. This is just the qth induced representation of $A^{(2)}$.[7] Since the above procedure can be recognized as a very simple method of induction and since induced matrices are a special class of invariant matrices, the property of involution is carried through for an arbitrary $n \times n$ matrix.[8]

In this paper we study the following:

(1) We show that the conditions on the 2×2 matrix $A^{(2)}$ such that the $[A^{(2)}]^m = kI$ are sufficient to make the qth induced matrix of $A^{(2)}$ obey the equation

$$[A_q^{(2)}]^m = k^q 1. \tag{1.4}$$

Of course, this is not surprising since induced matrices are invariant matrices.[7]

(2) We set up generating equations for the qth induced matrix of a 3×3 matrix $A^{(3)}$. In particular, if the matrix $A^{(3)}$ is involutional in the sense $[A^{(3)}]^m = kI$, the qth induced matrix $A_q^{(3)}$ satisfies the equation

$$[A_q^{(3)}]^m = k^q 1. \tag{1.5}$$

(3) The above procedure incidentally makes obvious the method of writing down the generating equation for the qth induced matrix of an arbitrary $n \times n$ matrix $A^{(n)}$. A particular case of interest is when $A^{(n)}$ is involutional.

(4) The special case of a 3×3 matrix $A^{(3)}$ satisfying $[A^{(3)}]^3 = 1$ is discussed in detail. It is shown that it can be expanded in the basis of the generalized Clifford algebra C_2^3 with coefficients which are the generalized hyperbolic functions.

(5) We calculate the eigenvalues of the matrix belonging to $GL(n)$ obtained through induction, and specialize it to the case of involutional matrices.

In Appendix A, we summarize the relevant information on the generalized Clifford algebra. In Appendix B we summarize the properties of hyperbolic and

trigonometric functions of order n. In Appendix C we demonstrate in a simple case the simplicity achieved in finding the arbitrary power of a matrix when it is expanded in the basis furnished by the roots of the unit matrix.

2. INVOLUTIONAL TRANSFORMATIONS OF $GL(2)$

The complete set of qth-degree polynomials in two variables x and y,

$$F_\nu(\mathbf{r}) = x^{q-\nu}y^\nu,$$
$$\mathbf{r} = (x, y), \quad \nu = 0, 1, \cdots, q, \tag{2.1}$$

is taken as the basis set. An element $R^{(2)}$ of $GL(2)$ is given by the general 2×2 matrix (nonsingular)

$$R^{(2)} = \begin{bmatrix} a & b \\ c & d \end{bmatrix} \in GL(2), \quad ad - bc \neq 0, \tag{2.2}$$

where a, b, c, and d are arbitrary parameters. The $(q+1)$-dimensional representation $R_q^{(2)}$ is furnished by the qth induced matrix of $R^{(2)}$ and is given by[6]

$$F_\nu(R^{(2)}\mathbf{r}) = (ax + by)^{q-\nu}(cx + dy)^\nu$$
$$= \sum_{\mu=0}^{q} [R_q^{(2)}]_{\nu\mu} F_\mu(\mathbf{r}). \tag{2.3}$$

The explicit form of $R_q^{(2)}$ is obtained by developing Eq. (2.3) in power series, and one gets

$$[R_q^{(2)}]_{\nu\mu} = a^{q-\mu-\nu}b^\mu c^\nu \sum_k \binom{\nu}{k}\binom{q-\nu}{\mu-k}\left(\frac{ad}{bc}\right)^k. \tag{2.4}$$

Now, an invariant matrix A_q of a matrix A is defined by the relation[7]

$$A_q B_q = (AB)_q, \tag{2.5}$$

where A_q is the matrix whose entries are polynomials in the elements of the matrix A, from which it easily follows that

$$[A_q]^m = [A^m]_q = k^q 1, \tag{2.6}$$

if

$$A^m = k1, \tag{2.7}$$

where A_q is of dimension $(q+1)$.

Therefore the conditions on the four parameters of the 2×2 matrix, in order that Eq. (2.7) is satisfied, automatically leave $A_q^{(2)}$ involutional. When $k = 1$, the involutional matrix $A^{(2)}$ involves only two parameters since in this case $a + d = 0$ and $bc = 1 - a^2$, and thus it can be expressed as

$$A^{(2)}(\theta) = \sigma_Z R(\theta), \tag{2.8}$$

where

$$\sigma_Z = \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix} \tag{2.9}$$

and $R(\theta)$ is the rotation matrix in two dimensions, with θ defined by

$$(bc)^{\frac{1}{2}} = \sin\theta. \tag{2.10}$$

3. GENERATING EQUATIONS FOR THE GENERAL INVOLUTIONAL MATRICES

By an exactly similar procedure as that utilized for the case of $m = 2$, we define the qth-degree homogeneous polynomials in three variables x, y, and z as

$$F_{\alpha_1,\alpha_2}(\mathbf{r}) = x^{q-\alpha_1-\alpha_2}y^{\alpha_1}z^{\alpha_2}, \tag{3.1}$$

where the nonnegative integers α_1 and α_2 obey

$$\alpha_1 + \alpha_2 \leq q. \tag{3.2}$$

The linear homogeneous transformation $R^{(3)}$ in three dimensions is given by a 3×3 matrix

$$R^{(3)} = \begin{bmatrix} a_{00} & a_{01} & a_{02} \\ a_{10} & a_{11} & a_{12} \\ a_{20} & a_{21} & a_{22} \end{bmatrix} \in GL(3), \tag{3.3}$$

where the a_{ij} are arbitrary parameters. The qth induced matrix of $R^{(3)}$ involving nine parameters is then simply given by the equation

$$F_{(\alpha_1\alpha_2)}(R^{(3)}\mathbf{r}) = (a_{00}x + a_{01}y + a_{02}z)^{q-\alpha_1-\alpha_2} \times (a_{10}x + a_{11}y + a_{12}z)^{\alpha_1} \times (a_{20}x + a_{21}y + a_{22}z)^{\alpha_2}$$
$$= \sum_{(\alpha_1'\alpha_2')} [R_q^{(3)}]_{(\alpha_1\alpha_2)(\alpha_1'\alpha_2')} F_{(\alpha_1'\alpha_2')}(\mathbf{r}),$$
$$r = (x, y, z), \tag{3.4}$$

where the matrix $R_q^{(3)}$ is labeled by the different partitions of the nonnegative integers (α_1', α_2') and (α_1, α_2) satisfying

$$\alpha_1' + \alpha_2' \leq q, \quad \alpha_1 + \alpha_2 \leq q. \tag{3.5}$$

Hence the dimension of R_q^3 is simply given by the number of solutions (α_1, α_2) of Eq. (3.5), which in this case is equal to $\binom{q+2}{2}$.

Obviously, $R_q^{(3)}$ reduces to $R^{(3)}$ when $q = 1$. For convenience, we can choose the partitions in decreasing order in α_1 for a given value of $\alpha_1 + \alpha_2$ and increasing order in $(\alpha_1 + \alpha_2)$ for labeling the matrix.

The simple power series expansion of Eq. (3.4)

yields an explicit expression for $R_q^{(3)}$, viz.,

$$
\begin{aligned}
[R_q^{(3)}]_{(\alpha_1\alpha_2)(\alpha_1'\alpha_2')} = {} & \left(\frac{a_{21}}{a_{20}}\right)^{\alpha_1'+\alpha_2'}\left(\frac{a_{22}}{a_{21}}\right)^{\alpha_2'} \\
& \times (a_{00})^{q-\alpha_1-\alpha_2}(a_{10})^{\alpha_1}(a_{20})^{\alpha_2} \\
& \times \sum_{\substack{\mu_1,\mu_2 \\ \nu_1,\nu_2}} \binom{q-\alpha_1-\alpha_2}{\mu_1}\binom{\alpha_1}{\nu_1} \\
& \times \binom{\alpha_2}{\alpha_1'+\alpha_2'-\mu_1-\nu_1}\binom{\mu_1}{\mu_2}\binom{\nu_1}{\nu_2} \\
& \times \binom{\alpha_1'+\alpha_2'-\mu_1-\nu_1}{\alpha_2'-\mu_2-\nu_2} \\
& \times \left(\frac{a_{01}a_{20}}{a_{00}a_{21}}\right)^{\mu_1}\left(\frac{a_{02}a_{21}}{a_{01}a_{22}}\right)^{\mu_2} \\
& \times \left(\frac{a_{11}a_{20}}{a_{10}a_{21}}\right)^{\nu_1}\left(\frac{a_{12}a_{21}}{a_{11}a_{22}}\right)^{\nu_2}. \qquad (3.6)
\end{aligned}
$$

This may possibly be related to the Lauricella functions.[9]

The above procedure of generating the induced matrix can be easily generalized to the case of an arbitrary $n \times n$ matrix $R^{(n)}$ given by

$$
R^{(n)} = \begin{bmatrix} a_{00} & a_{01} & \cdots & a_{0n-1} \\ a_{10} & a_{11} & \cdots & a_{1n-1} \\ \cdot & \cdot & \cdots & \cdot \\ \cdot & \cdot & \cdots & \cdot \\ \cdot & \cdot & \cdots & \cdot \\ a_{n-1,0} & a_{n-1,1} & \cdots & a_{n-1,n-1} \end{bmatrix} \in GL(n). \qquad (3.7)
$$

In this case we define the qth-degree polynomials $F_{(\alpha_1,\cdots,\alpha_{n-1})}(\mathbf{r})$ in n variables $x_1, \cdots, x_n$,

$$
\begin{aligned}
F_{(\alpha_1,\cdots,\alpha_{n-1})}(\mathbf{r}) &= x_1^{q-\sum_{i=1}^{n-1}\alpha_i} x_2^{\alpha_1}, \cdots, x_n^{\alpha_{n-1}}, \\
\mathbf{r} &= (x_1, \cdots, x_n), \qquad (3.8)
\end{aligned}
$$

with the nonnegative integers α_i satisfying the partition equation

$$
\sum_{i=1}^{n-1} \alpha_i \leq q. \qquad (3.9)
$$

The qth induced representation of $R^{(n)}$ is given by the matrix $R_q^{(n)}$ defined by

$$
\begin{aligned}
& F_{\alpha_1,\cdots,\alpha_{n-1}}(R^{(n)}\mathbf{r}) \\
& = (a_{00}x_1 + \cdots + a_{0n-1}x_n)^{q-\sum_{i=1}^{n-1}\alpha_i} \prod_{j=1}^{n-1}\left(\sum_{k=0}^{n-1} a_{jk}x_{k+1}\right)^{\alpha_j} \\
& = \sum_{(\alpha_1',\cdots,\alpha_{n-1}')} [R_q^{(n)}]_{(\alpha_1,\cdots,\alpha_{n-1})(\alpha_1',\cdots,\alpha_{n-1}')} F_{(\alpha_1',\cdots,\alpha_{n-1}')}(\mathbf{r}), \qquad (3.10)
\end{aligned}
$$

where the matrix is labeled by the distinct partitions given by Eq. (3.9). We can choose them in the decreasing order $(\alpha_1, \cdots, \alpha_{n-1})$ for a given value of $(\alpha_1 + \cdots + \alpha_{n-1}) \leq q$.

The dimension of $R_q^{(n)}$ is just given by the number of solutions of the partitions equation (3.9), which is simply equal to $\binom{n+q-1}{q} = \binom{n+q-1}{n-1}$.

Let us now specialize the method of induction to the case of involutional matrices satisfying the equation

$$
[R^{(n)}]^m = kI. \qquad (3.11)
$$

As in the case of (2×2) matrix, the conditions on $R^{(n)}$, so that it satisfies Eq. (3.11), leave its qth induced representation $R_q^{(n)}$ to obey

$$
[R_q^{(n)}]^m = k^q I. \qquad (3.12)
$$

This follows directly from the property of induced matrices, which form a special case of invariant matrices satisfying Eq. (2.5). The conditions on $R^{(n)}$ implied by Eq. (3.11), when $m = n$, are simply given by the characteristic equation of $R^{(n)}$:

$$
\operatorname{Tr} R^{(n)} = \operatorname{Tr} [R^{(n)}]^2 = \cdots = \operatorname{Tr} [R^{(n)}]^{n-1} = 0,
$$

and

$$
\det R^{(n)} = (-1)^n k. \qquad (3.13)
$$

Let us consider the special case of a 3×3 matrix satisfying the equation

$$
[A^{(3)}]^3 = 1. \qquad (3.14)
$$

The eigenvalues of $A^{(3)}$ are then given by the cube roots of unity $(1, \omega, \omega^2)$. As in the case of $A^{(2)}$, $A^{(3)}$ can be reduced to the form

$$
F^{(3)}(\theta) = VA^{(3)}V^{-1} = \begin{bmatrix} f_1^{(3)} & \omega f_2^{(3)} & f_3^{(3)} \\ f_3^{(3)} & \omega f_1^{(3)} & f_2^{(3)} \\ \omega^2 f_2^{(3)} & f_3^{(3)} & \omega^2 f_1^{(3)} \end{bmatrix}, \qquad (3.15)
$$

where the $f_i^{(3)}$ are the generalized hyperbolic functions of order three with argument $(\lambda\theta)$, with $\lambda = \exp(\frac{1}{3}\pi i) = \omega^{\frac{1}{2}}$, ω being a primitive cube root of unity. It is inessential to compute the matrix V whose existence can be inferred from the fact that both $A^{(3)}$ and $F^{(3)}$ are nonsingular and satisfy Eq. (3.14). The f's are functions of the entries of the matrix $A^{(3)}$. They satisfy the determinantal condition[10]

$$
\begin{vmatrix} f_1^{(3)} & f_2^{(3)} & f_3^{(3)} \\ f_3^{(3)} & f_1^{(3)} & f_2^{(3)} \\ f_2^{(3)} & f_3^{(3)} & f_1^{(3)} \end{vmatrix} = 1. \qquad (3.16)
$$

The $f_i^{(3)}$ are related to the trigonometric functions of order three $k_i^{(3)}$ (see Appendix B) through the relation

$$
k_i^{(3)}(\theta) = \lambda^{(1-i)} f_i^{(3)}(\lambda\theta). \qquad (3.17)
$$

$F^{(3)}(\theta)$ can be expressed as

$$
F^{(3)}(\theta) = B^{(3)}R^{(3)}(\theta), \qquad (3.18)
$$

where

$$B^{(3)} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & \omega & 0 \\ 0 & 0 & \omega^2 \end{bmatrix} \tag{3.19}$$

and $R^{(3)}(\theta)$ is the matrix

$$R^{(3)}(\theta) = \begin{bmatrix} f_1 & \omega f_2 & f_3 \\ \omega^2 f_3 & f_1 & \omega^2 f_2 \\ f_2 & \omega f_3 & f_1 \end{bmatrix}. \tag{3.20}$$

The interesting point is that $R^{(3)}(\theta)$ can be expressed as

$$R^{(3)}(\theta) = \sum_{i=1}^{3} f_i^{(3)}(\lambda\theta) Q_3^{i-1} \tag{3.21}$$

$$= \sum_{i=1}^{3} \lambda^{i-1} k_i^{(3)}(\theta) Q_3^{i-1}, \tag{3.22}$$

where the matrix

$$Q_3 = \begin{bmatrix} 0 & \omega & 0 \\ 0 & 0 & \omega^2 \\ 1 & 0 & 0 \end{bmatrix} \tag{3.23}$$

is a base element of the generalized Clifford algebra C_2^3 (see Appendix A). The determinantal condition (3.16) can also be written as

$$\det \sum_{i=1}^{3} f_i^{(3)}(\lambda\theta) P_3^{i-1} = \det \sum_{i=1}^{3} k_i^{(3)}(\theta) \lambda^{i-1} P_3^{i-1} = 1, \tag{3.24}$$

where the matrix

$$P_3 = \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 1 & 0 & 0 \end{bmatrix} \tag{3.25}$$

is the other base element of C_3^2.

The above discussion can now be carried for an arbitrary $(n \times n)$ involutional matrix

$$[A^{(n)}]^n = 1, \tag{3.26}$$

which can be transformed to the form

$$V^{(n)} A^{(n)} V^{-(n)} = F^{(n)}(\theta) = B^{(n)} R^{(n)}(\theta), \tag{3.27}$$

where

$$B^{(n)} = \begin{bmatrix} 1 & 0 & 0 & 0 & \cdots & 0 \\ 0 & \omega & 0 & 0 & \cdots & 0 \\ . & . & . & . & \cdots & . \\ 0 & 0 & . & . & \cdots & \omega^{n-1} \end{bmatrix} \tag{3.28}$$

and[11]

$$R^{(n)}(\theta) = \sum_{i=1}^{n} f_i^{(n)}(\lambda\theta) Q_n^{i-1}$$

$$= \sum_{i=1}^{n} \lambda^{i-1} k_i^{(n)}(\theta) Q_n^{i-1}, \quad \lambda = \exp(n^{-1}\pi i), \tag{3.29}$$

where the matrix

$$Q_n = \zeta \begin{bmatrix} 0 & \omega & 0 & \cdots & 0 \\ 0 & 0 & \omega^2 & \cdots & 0 \\ 0 & 0 & . & \cdots & \omega^{n-1} \\ 1 & 0 & 0 & \cdots & 0 \end{bmatrix},$$

$$\zeta = 1 \text{ for } n \text{ odd}$$

$$= \lambda \text{ for } n \text{ even}, \tag{3.30}$$

is an element of the generalized Clifford algebra C_2^n. The determinantal condition on the hyperbolic and trigonometric functions of order n is simply given by

$$\det \sum_{i=1}^{n} f_i^{(n)}(\lambda\theta) P_n^{i-1} = \det \sum_{i=1}^{n} \lambda^{i-1} k_i^{(n)}(\theta) P_n^{i-1} = 1, \tag{3.31}$$

where the matrix

$$P_n = \begin{bmatrix} 0 & 1 & 0 & \cdots & 0 \\ 0 & 0 & 1 & \cdots & 0 \\ 0 & 0 & 0 & \cdots & 1 \\ 1 & 0 & 0 & \cdots & 0 \end{bmatrix} \tag{3.32}$$

is the other base element of C_2^n. The f's are functions of the entries of the matrix $A^{(n)}$, and the explicit relation is of little concern to us. The existence of $V^{(n)}$ is again guaranteed by the fact that $A^{(n)}$ and $F^{(n)}(\theta)$ are both nonsingular and satisfy Eq. (3.26).

4. EIGENVALUES OF $R_q^{(n)}$ AND $A_q^{(n)}$

In this section we first calculate the eigenvalues of the qth induced matrix $R_q^{(n)}$ of the matrix $R^{(n)}$ given by Eq. (3.10) and specialize to the case when $R_q^{(n)}$ is involutional. The calculation is based on the simple theorem that if the matrix $R^{(n)}$ is triangular, then its induced matrix $R_q^{(n)}$ is also triangular in shape similar to $R^{(n)}$. This theorem has been proved by Kim[6] for $n = 2$, and it is true even in the general case. Consider for example the case of $n = 3$. If $R^{(3)}$ has the form

$$R_t^{(3)} = \begin{bmatrix} a_{00} & 0 & 0 \\ a_{10} & a_{11} & 0 \\ a_{20} & a_{21} & a_{22} \end{bmatrix}, \tag{4.1}$$

it follows from Eq. (3.4) that

$$[R_q^{(3)}]_{\alpha\alpha'} = 0,$$

unless $\alpha_2' \leq \alpha_2$ and $\alpha_1' + \alpha_2' \leq \alpha_1 + \alpha_2$, where $\alpha = (\alpha_1, \alpha_2)$ and $\alpha' = (\alpha_1', \alpha_2')$. These are simply the conditions for the matrix $R_q^{(3)}$ to be triangular in shape similar to $R^{(3)}$. It is not hard to prove the same result for any n.

In fact, it follows directly from Eq. (3.10) that if the matrix R^n is triangular, then, since $a_{ij} = 0$, $i < j$,

we have $[R_q^n]_{\alpha\alpha'} = 0$ unless

$$\sum_{i=k}^{n-1}\alpha_i \geq \sum_{i=k}^{n-1}\alpha_i', \quad k = 1, \cdots, n-1. \tag{4.2}$$

These are simply the conditions for $R_q^{(n)}$ to be triangular and similar in shape to $R_i^{(n)}$. Equation (4.2) incidentally suggests a more convenient labeling of $R_q^{(n)}$ by $(\beta_1, \cdots, \beta_{n-1})$ satisfying

$$0 \leq \beta_{n-1} \leq \cdots \leq \beta_1 \leq q,$$
$$\beta_1 + \beta_2 + \cdots + \beta_{n-1} \leq (n-1)q,$$

where

$$\beta_j = \sum_{i=j}^{n-1}\alpha_i, \quad j = 1, \cdots, n-1. \tag{4.3}$$

The generating equation (3.10) for the induced matrix $R_q^{(n)}$ then simply becomes

$$\prod_{j=0}^{n-1}\left(\sum_{k=0}^{n-1}a_{jk}x_{k+1}\right)^{\beta_j-\beta_{j+1}} = \sum_{\beta'}[R_q^{(n)}]_{\beta\beta'}F_{\beta'}(\mathbf{r}) \tag{4.4}$$

with

$$\beta_0 = q, \quad \beta_n = 0. \tag{4.5}$$

Equations (3.10) and (4.4) are completely equivalent.

Now it is always possible to transform the matrix $R^{(3)}$ into the triangular matrix $R^{(3)T}$,

$$R^{(3)T} = \begin{bmatrix} \epsilon_1 & 0 & 0 \\ \xi_1 & \epsilon_2 & 0 \\ \xi_2 & \xi_3 & \epsilon_3 \end{bmatrix}, \tag{4.6}$$

through a suitable unitary transformation. Here the ξ's are constants, and the ϵ's are the eigenvalues of $R^{(3)}$. Substituting Eq. (4.6) in Eq. (3.6), we obtain

$$\begin{aligned}[R_q^{(3)T}]_{\alpha\alpha'} &= \epsilon_1^{q-\alpha_1-\alpha_2}\epsilon_3^{\alpha_2'} \\ &\quad \times \sum_{\nu}\binom{\alpha_1}{\nu}\binom{\alpha_2}{\alpha_1'+\alpha_2'-\nu}\binom{\alpha_1'+\alpha_2'-\nu}{\alpha_2'} \\ &\quad \times (\epsilon_2)^{\nu}(\xi_1)^{\alpha_1-\nu}(\xi_2)^{\alpha_2-\alpha_1'-\alpha_2'+\nu}(\xi_3)^{\alpha_1'-\nu}, \\ &\qquad\qquad \xi_1, \xi_2, \xi_3 \neq 0, \\ &= \epsilon_1^{q-\alpha_1-\alpha_2}(\epsilon_2)^{\alpha_1}(\epsilon_3)^{\alpha_2}\delta_{\alpha_1\alpha_1'}\delta_{\alpha_2\alpha_2'}, \\ &\qquad\qquad \xi_1 = \xi_2 = \xi_3 = 0. \end{aligned} \tag{4.7}$$

The eigenvalues of $R_q^{(3)}$ are then given by

$$\epsilon_1^{q-\alpha_1-\alpha_2}\epsilon_2^{\alpha_1}\epsilon_3^{\alpha_2}, \quad \alpha_1 + \alpha_2 \leq q. \tag{4.8}$$

The determinant of $R_q^{(3)}$ is given by

$$\begin{aligned}\det R_q^{(3)} &= \prod_{\substack{\alpha_1,\alpha_2 \\ \alpha_1+\alpha_2\leq q}}\epsilon_1^{q-\alpha_1-\alpha_2}\epsilon_2^{\alpha_1}\epsilon_3^{\alpha_2} \\ &= (\epsilon_1\epsilon_2\epsilon_3)^{\binom{q+2}{3}} = \Delta^{\binom{q+2}{3}}, \end{aligned} \tag{4.9}$$

where Δ denotes the determinant of $R^{(3)}$

The trace of $R_q^{(3)}$ is given by

$$\begin{aligned}\operatorname{Tr} R_q^{(3)} &= 1/[(\epsilon_1-\epsilon_2)(\epsilon_2-\epsilon_3)(\epsilon_1-\epsilon_3)] \\ &\quad \times [\epsilon_1\epsilon_2(\epsilon_1^{q+1}-\epsilon_2^{q+1}) + \epsilon_2\epsilon_3(\epsilon_2^{q+1}-\epsilon_3^{q+1}) \\ &\quad + \epsilon_3\epsilon_1(\epsilon_3^{q+1}-\epsilon_1^{q+1})], \quad \epsilon_1 \neq \epsilon_2 \neq \epsilon_3, \end{aligned} \tag{4.10}$$

$$= \binom{q+2}{2}\epsilon^q, \quad \epsilon_1 = \epsilon_2 = \epsilon_3 = \epsilon. \tag{4.11}$$

The above formula can be immediately generalized to yield

$$\det R_q^{(n)} = (\Delta)^{\binom{q+n-1}{n}}, \tag{4.12}$$

where

$$\Delta = \epsilon_1\epsilon_2\cdots\epsilon_n \tag{4.13}$$

is the determinant of $R^{(n)}$. Further, we have

$$\begin{aligned}\operatorname{Tr} R_q^{(n)} &= \prod_{\substack{i<j \\ i,j=1,\cdots,n}}(\epsilon_i-\epsilon_j)^{-1}\sum_{\substack{k\neq l \\ k,l \text{ cyclic} \\ k,l=1,\cdots,n}}\epsilon_k\epsilon_l[\epsilon_k^{q+1}-\epsilon_l^{q+1}], \\ &\qquad\qquad \epsilon_1 \neq \epsilon_2 \neq \cdots \neq \epsilon_n, \\ &= \binom{q+n-1}{n-1}\epsilon^q, \quad \epsilon_1 = \epsilon_2 = \cdots = \epsilon_n = \epsilon. \end{aligned} \tag{4.14}$$

The eigenvalues of $R_q^{(n)}$ are given by

$$\epsilon_1^{q-\sum_{i=1}^{n-1}\alpha_i}\epsilon_2^{\alpha_1}\epsilon_3^{\alpha_2}\epsilon_n^{\alpha_{n-1}}, \quad \sum_{i=1}^{n-1}\alpha_i \leq q. \tag{4.15}$$

Another interesting property of $R_q^{(n)}$ which can be easily verified from Eq. (3.10) is that

$$\sum_{i,j=0}^{n-1}a_{ij}\frac{\partial R_q^{(n)}}{\partial a_{ij}} = qR_q^{(n)}. \tag{4.16}$$

The above discussion can now be specialized to the case of the general involutional $n \times n$ matrix $A^{(n)}$. The eigenvalues of $A^{(n)}$ are given by

$$\epsilon_1 = \epsilon, \epsilon_2 = \omega\epsilon, \epsilon_3 = \omega^2\epsilon_2\cdots\epsilon_n = \omega^{n-1}\epsilon,$$
$$\omega^n = 1. \tag{4.17}$$

In this case we have

$$\operatorname{Tr} A_q^{(n)} = \frac{(-1)^n\epsilon^q}{n}(1 + \omega^q + \omega^{2q} + \cdots + \omega^{(n-1)q}), \tag{4.18}$$

so that

$$\begin{aligned}\operatorname{Tr} A_q^{(n)} &= 0 \quad \text{for} \quad q \neq 0 \bmod n \\ &= (-1)^n\epsilon^q \quad \text{for} \quad q = 0 \bmod n. \end{aligned} \tag{4.19}$$

The determinant of $A_q^{(n)}$ is given by

$$\det A_q^{(n)} = [\epsilon^n\omega^{\binom{n}{2}}]^{\binom{q+n-1}{n}}. \tag{4.20}$$

ACKNOWLEDGMENTS

The authors are grateful to Professor A. Ramakrishnan for very fruitful discussions regarding

the roots of the unit matrix and Professor A. O. Morris for drawing our attention to induced matrices. Thanks are also due to Dr. R. Vasudevan for certain remarks. One of the authors (T. S. S.) is grateful to Professor K. Dietz, Professor H. Rollnik, and other members of the Physikalisches Institut for their hospitality at the institute. Another author (N. B. M.) is grateful to the C.S.I.R., Government of India, for the award of a Junior Fellowship.

APPENDIX A

We summarize here the relevant details of the generalized Clifford algebra.[2-4] The equation

$$\sum_{i=1}^{n} x_i^m = \left(\sum_{i=1}^{n} \alpha_i x_i\right)^m \tag{A1}$$

is satisfied if α's obey the relations

$$\alpha_i \alpha_j = \omega \alpha_j \alpha_i, \quad i < j, i, j = 1, \cdots, n,$$
$$\omega^m = 1. \tag{A2}$$

The set of elements defined by

$$\prod_{i=1}^{m} \alpha_i^{\rho_i}, \tag{A3}$$

where the integers ρ_i satisfy

$$0 \leq \rho_i \leq m - 1, \tag{A4}$$

is linearly independent. They are m^n in number. Obviously they form a vector space of dimension m^n, and with the product defined by Eq. (A2) they form an associative algebra called the generalized Clifford algebra C_n^m. The case when $m = 2$ can be realized to be the Dirac Clifford algebra. The matrix representation of α's has been obtained by using the Dirac procedure by Morinaga and Nono[2] and has also been obtained by Ramakrishnan, Santhanam, and Chandrasekaran[12] by using vector space methods. The results are: C_n^m for $n = 2\nu$ has a faithful representation by the matrix ring $m^\nu \times m^\nu$; when n is odd, it has again the matrix representation in terms of $(m^\nu \times m^\nu)$-dimensional matrices, which, however, breaks up into m sets of inequivalent matrix rings $m^\nu \times m^\nu$. That is, if the set $\{\beta\}$ furnishes a representation of dimension $m^\nu \times m^\nu$, then $\omega^i\{\beta\}$, $i = 1, \cdots, m - 1$ also furnish inequivalent representations of the same dimension, ω being a primitive mth root of unity. The case when $m = 2$ is, of course, very well known.[13]

APPENDIX B

We summarize here some general properties of the trigonometric and hyperbolic functions of order n.[10]

The n functions

$$f_i = \frac{1}{n} \sum_{m=1}^{n} \omega^{(1-i)m} \exp(\omega^m x), \quad i = 1, \cdots, n,$$
$$\omega = \exp\left(\frac{2\pi i}{n}\right), \tag{B1}$$

are called the hyperbolic functions of order n. The f_i satisfy the differential equation

$$\left(\frac{d^n}{dx^n} - 1\right) y = 0. \tag{B2}$$

$f_1, \cdots, f_n$ form a linearly independent set of solutions of Eq. (B2) and their Wronskian is equal to unity. From the definition of f_i follows that

$$\exp(\omega^m x) = \sum_{i=1}^{n} \omega^{(1-i)m} f_i(x, n), \quad m \text{ integer}. \tag{B3}$$

From (B3) it follows that

$$\prod_{m=1}^{n} \left(\sum_{i=1}^{n} \omega^{(1-i)m} f_i(x, n)\right) = 1. \tag{B4}$$

Equation (B4) can also be written as

$$\det \sum_{i=1}^{n} f_i P^{i-1} = 1, \tag{B5}$$

where the permutation matrix

$$P = \begin{bmatrix} 0 & 1 & 0 & \cdots & 0 \\ 0 & 0 & 1 & \cdots & 0 \\ 0 & 0 & 0 & \cdots & 1 \\ 1 & 0 & 0 & 0 & 0 \end{bmatrix} \tag{B6}$$

is a base element of C_2^n.

The functions

$$k_i(x, n) = \lambda^{1-i} f_i(\lambda x, n), \quad i = 1, \cdots, n,$$
$$\lambda = \exp(\pi i/n), \tag{B7}$$

are called the trigonometric functions of order n. They are the solutions of the differential equation

$$\left(\frac{d^n}{dx^n} + 1\right) y = 0. \tag{B8}$$

From (B7) it is clear that

$$k_i(x, n) = \frac{1}{n} \sum_{m=1}^{n} \lambda^{(1-i)(2m+1)} \exp(\lambda^{2m+1} x) \tag{B9}$$

and

$$\prod_{m=1}^{n} \left(\sum_{i=1}^{n} \lambda^{(1-i)(2m+1)} K_i(x)\right) = \det \sum_{i=1}^{n} \lambda^{(i-1)} K_i Q^{i-1} = 1. \tag{B10}$$

APPENDIX C

We demonstrate the use of the expansion of a matrix in the basis of the roots of unit matrix to find its arbitrary power. The problem of finding the arbitrary power of an $n \times n$ matrix has already attracted the attention of many people.[14] We believe that the expansion of a matrix in terms of the roots of the unit matrix will simplify the problem very much. Since we have the results only for the case of a 2×2 matrix, which is very well known, we content ourselves by just giving the results. Any 2×2 matrix X can be uniquely expanded as

$$X = l_0 1 + \mathbf{l} \cdot \boldsymbol{\sigma}, \tag{C1}$$

where the σ's are the Pauli matrices forming the algebra c_2^2 along with the unit matrix. If the matrix X has the form

$$X = \begin{bmatrix} a & b \\ c & d \end{bmatrix},$$

then

$$\begin{aligned} l_0 &= \tfrac{1}{2}(a + p), \quad l_3 = \tfrac{1}{2}(a - d), \\ l_1 &= \tfrac{1}{2}(b + c), \quad l_2 = \tfrac{1}{2}[i(b - c)]. \end{aligned} \tag{C2}$$

Then it is easy to see that

$$\begin{aligned} X^m = {} & \tfrac{1}{2}[(l_0 + l)^m + (l_0 - l)^m] \\ & + \frac{1}{2}\frac{(\mathbf{l} \cdot \sigma)}{l}[(l_0 + l)^m - (l_0 - l)^m], \end{aligned} \tag{C3}$$

where

$$l = +(l_1^2 + l_2^2 + l_3^2)^{\frac{1}{2}}. \tag{C4}$$

Equation (C3) can also be written as

$$\begin{aligned} & X^m = U_m(p, q)X + qU_{m-1}(p, q)1, \\ & p = (a + d) = \operatorname{Tr} X, \quad \text{and} \quad q = ad - bc = \det X, \end{aligned} \tag{C5}$$

where the U's are the Luca's polynomials given by

$$\begin{aligned} U_m(p, q) = {} & \frac{1}{2^m(p^2 - 4q)^{\frac{1}{2}}} \\ & \times [[p + (p^2 - 4q)^{\frac{1}{2}}]^m \\ & - [p - (p^2 - 4q)^{\frac{1}{2}}]^m]. \end{aligned}$$

Of course, there are methods of Sylvester using the explicit eigenvalues of X and the method of using the characteristic equation of X. But we hope that the expansion in terms of the roots of unit matrix can be much simpler, as in the case of $m = 2$ demonstrated above. The simple reason is that the (nontrivial) roots of the unit matrix are traceless matrices, and hence their characteristic equation is much simpler. The details for $m > 2$ will be published elsewhere.

* On leave of absence from Matscience, The Institute of Mathematical Sciences, Madras 20, India.

[1] A. Ramakrishnan *et al.*, J. Math. Anal. Appl. **27**, 164 (1969); L. A. Pipes, J. Franklin Inst. **287**, 285 (1969); S. K. Kim, J. Math. Phys. **9**, 1705 (1968); M. E. Fisher, Phys. Rev. **113**, 969 (1959); see also A. Deepak *et al.*, Intern. J. Quant. Chem. **3**, 445 (1969).

[2] K. Morinaga and T. Nono, J. Sci. Hiroshima Univ. **A6**, 13 (1952).

[3] K. Yamazaki, J. Fac. Sci. Univ. Tokyo, Sec. 1, **10**, 147 (1964).

[4] A. O. Morris, Quart. J. Math. (Oxford) (2) **18**, 7 (1967); **19**, 289 (1968).

[5] A. Ramakrishnan, J. Math. Anal. Appl. **20**, 9 (1967).

[6] S. K. Kim, J. Math. Phys. **10**, 1225 (1969).

[7] D. E. Littlewood, *The Theory of Group Characters and Matrix Representations of Groups* (Oxford U.P., Oxford, 1958), p. 178.

[8] It has been pointed out to us by Professor A. Ramakrishnan that the method of induction can be related to taking the direct product of helicity matrices defined by him. See A. Ramakrishnan, J. Math. Anal. Appl. **26**, 88 (1969).

[9] J. Slater, *Generalized Hypergeometric Functions* (Cambridge U.P., Cambridge, 1966), p. 227.

[10] A. Erdélyi, *Higher Transcendental Functions* (McGraw-Hill, New York, 1955), Vol. III, pp. 212–17.

[11] It has been noted by Professor A. Ramakrishnan (private communication) that if a matrix $T(x)$ has the form $T(x) = Me^{Nx}$, with the matrices M and N satisfying the relation $MN = \omega NM$, $M^m = N^m = I$, then it follows that $[T(x)]^m = I$.

[12] A. Ramakrishnan, T. S. Santhanam, and P. S. Chandrasekaran, J. Math. Phys. Sci. (Madras) **3**, 307 (1969).

[13] See, for example, H. Boerner, *Representation of Groups* (North-Holland, Amsterdam, 1963), Chap. 8.

[14] A. Herpin, Compt. Rend. Acad. Sci. (Paris) **225**, 17 (1947). Recently this problem has been studied by R. Bakarat, J. Math. & Phys. **43**, 332 (1964); R. Bakarat and E. Baumann, J. Math. Phys. **10**, 1474 (1969).

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 12, NUMBER 3 MARCH 1971

Labeling States and Constructing Matrix Representations of F_4

J. PATERA
Centre de recherches mathématiques, Université de Montréal, Montréal, Canada

(Received 10 August 1970)

A method is developed for construction of a complete basis for a class of finite-dimensional irreducible representations of the simple Lie algebra F_4 and for calculation of matrix elements of F_4-generators relative to that basis. The algebra F_4 is embedded into A_{25}, and results of Gel'fand and Tsetlin concerning that algebra are used.

I. INTRODUCTION

The present paper is the second in a series of three in which a method is developed for construction of a complete basis in any finite-dimensional representation space of any exceptional simple Lie algebra and for calculation of matrix elements of representations relative to that basis. In the first paper of the series[1] (hereafter referred to as I) the irreducible representations (IR's) of the algebra G_2 were considered, and the general idea of the whole method was outlined. Section II of I contains several conventions and should be considered an integral part of the present paper as well. In what follows we refer to equation $(\cdots)$ of I as $(\mathrm{I}\cdots)$.

Here we study the problem for a class of IR's of F_4. The algebra F_4 is embedded into an algebra A_{25}, and results of Gel'fand and Tsetlin[2] concerning that algebra are used. This is precisely the way we proceeded in I. Unlike the case of G_2, however, an additional difficulty arises here.[3] A single embedding of F_4 into one algebra A_n does not suffice to treat all IR's of F_4. One would also need, for instance, an embedding $F_4 \subset A_{51}$ and others. The rather high rank of the containing algebra A_n in these cases necessitates a modification of our approach to the problem. Therefore, the IR's of F_4 not included in the present work [cf. condition (6) below], as well as all the IR's of algebras E_6, E_7, and E_8, will be considered elsewhere in a way which is, in principle, the same but which in some situations is not as conveniently applicable.

In Sec. II, the embedding $F_4 \subset A_{25}$ is described. Section III contains the state labeling lemma and the method for calculating matrix elements of F_4-generators. An example of matrix element calculation is shown in Sec. IV. The last section is devoted to some observations about the method.

II. EMBEDDING OF F_4 INTO A_{25}

First we assume all the conventions and notations of Sec. II of I. We introduce a basis $\{\tau_i\}$, $i = 1, 2, 3, 4$, in the root space $R(F_4)$ by

$$\begin{aligned}
\tau_1 &= 2\beta_1 + 3\beta_2 + 4\beta_3 + 2\beta_4, & \beta_1 &= 2\tau_1 - \tau_2,\\
\tau_2 &= 3\beta_1 + 6\beta_2 + 8\beta_3 + 4\beta_4, & \beta_2 &= -\tau_1 + 2\tau_2 - 2\tau_3,\\
\tau_3 &= 2\beta_1 + 4\beta_2 + 6\beta_3 + 3\beta_4, & \beta_3 &= -\tau_2 + 2\tau_3 - \tau_4,\\
\tau_4 &= \beta_1 + 2\beta_2 + 3\beta_3 + 2\beta_4, & \beta_4 &= -\tau_3 + 2\tau_4,
\end{aligned} \tag{1}$$

where the vectors β_i are simple roots of F_4. Equations (1) constitute a particular case of (11).

The embedding $F_4 \subset A_{25}$ is specified by a projection f^* of the 25-dimensional real vector space spanned by simple roots of A_{25} onto the 4-dimensional root space of F_4.

According to a general method,[4] the projection f^* is found when the weights

$$\begin{aligned}
&M_1 = \tau_4, && M_2 = \tau_3 - \tau_4, && M_3 = \tau_2 - \tau_3,\\
&M_4 = \tau_1 - \tau_2 + \tau_3, && M_5 = \tau_1 - \tau_3 + \tau_4, && M_6 = -\tau_1 + \tau_3,\\
&M_7 = \tau_1 - \tau_4, && M_8 = -\tau_1 + \tau_2 - \tau_3 + \tau_4, && M_9 = -\tau_1 + \tau_2 - \tau_4,\\
&M_{10} = -\tau_2 + \tau_3 + \tau_4, && M_{11} = -\tau_2 + 2\tau_3 - \tau_4, && M_{12} = -\tau_3 + 2\tau_4,\\
&M_{13} = M_{14} = 0,
\end{aligned} \tag{2}$$

and

$$M_{13+k} = -M_{13-k+1} \quad \text{for} \quad k = 1, 2, \cdots, 13,$$

of the lowest-dimensional representation of F_4 are numbered as in (2) and inserted into Table I of Ref. 4. From that table, one has

$$\begin{aligned} f^*(\nu_1) &= \tau_4, & f^*(\nu_6) &= \tau_1 + \tau_3 + \tau_4, \\ f^*(\nu_2) &= \tfrac{1}{3}f^*(\nu_{11}) = \tau_3, & f^*(\nu_7) &= 2\tau_1 + \tau_3, \\ f^*(\nu_3) &= \tfrac{1}{2}f^*(\nu_9) = \tau_2, & f^*(\nu_8) &= \tau_1 + \tau_2 + \tau_4, \\ f^*(\nu_4) &= \tau_1 + \tau_3, & f^*(\nu_{10}) &= \tau_2 + \tau_3 + \tau_4, \\ f^*(\nu_5) &= 2\tau_1 + \tau_4, & f^*(\nu_{12}) &= f^*(\nu_{13}) = 2\tau_3 + 2\tau_4 \end{aligned} \tag{3}$$

and

$$f^*(\nu_{13+k}) = f^*(\nu_{13-k}) \quad \text{for} \quad k = 1, 2, \cdots, 12.$$

Consequently, a vector $M = \sum_{k=1}^{25} b_k \nu_k$ of the root space of A_{25} becomes, after the projection f^*,

$$\begin{aligned} f^*(M) = N = {} & \tau_1(b_4 + 2b_5 + b_6 + 2b_7 + b_8 + b_{18} + 2b_{19} + b_{20} + 2b_{21} + b_{22}) \\ & + \tau_2(b_3 + b_8 + 2b_9 + b_{10} + b_{16} + 2b_{17} + b_{18} + b_{23}) \\ & + \tau_3(b_2 + b_4 + b_6 + b_7 + b_{10} + 3b_{11} + 2b_{12} + 2b_{13} + 2b_{14} \\ & + 3b_{15} + b_{16} + b_{19} + b_{20} + b_{22} + b_{24}) \\ & + \tau_4(b_1 + b_5 + b_6 + b_8 + b_{10} + 2b_{12} + 2b_{13} + 2b_{14} + b_{16} + b_{18} + b_{20} + b_{21} + b_{25}). \end{aligned} \tag{4}$$

Recalling Theorem 0.11 of Ref. 5, one finds that a Gel'fand pattern (g) [cf. (I3), where $n = 25$], which is a vector of a representation space $R(\Phi(A_{25}))$ and which has its A_{25}-weight given by (I5), also has a well-defined F_4-weight. It is obtained when (I5) is inserted into (4). Because $F_4 \subset A_{25}$, one has $R(\Phi(A_{25})) \equiv R(\Phi(F_4))$, where $\Phi(F_4)$ is the representation of F_4 (in general, reducible) obtained when the IR $\Phi(A_{25})$ is restricted to the subalgebra F_4 of A_{25}.

Suppose we are given an IR $\Psi(F_4)$ and we want to construct a basis in its representation space $R(\Psi(F_4))$. For that, we use the Gel'fand patterns of an IR $\Phi(A_{25})$. Obviously, this is possible only if $R(\Psi) \subseteq R(\Phi)$, which implies

$$\Psi(F_4) \subseteq \Phi(F_4) \subset \Phi(A_{25}). \tag{5}$$

Because the latter inclusion may hold for several $\Phi(A_{25})$, it is natural to choose $\Phi(A_{25})$ as "small" as possible. More precisely, our choice is made by

Lemma 1: Let $\Psi(F_4)$ and $\Phi(A_{25})$ be IR's of F_4 and A_{25} with the highest weights

$$N_h = \sum_{k=1}^{4} a_k \tau_k, \quad a_1 \leq a_3, \tag{6}$$

and

$$M_h = a_4\nu_1 + (a_3 - a_1)\nu_2 + a_2\nu_3 + a_1\nu_4, \tag{7}$$

respectively. Then:

(a) $\Phi(A_{25}) \supset \Psi(F_4)$;

(b) if also $\Omega(A_{25}) \supset \Psi(F_4)$, then the dimension of $\Omega(A_{25})$ is not lower than that of $\Phi(A_{25})$;

(c) patterns (I3) form a basis in $R(\Phi)$ provided that

$$\begin{aligned} m_{1,26} &= a_2 + a_3 + a_4, & m_{2,26} &= a_2 + a_3, \\ m_{3,26} &= a_2 + a_1, & m_{4,26} &= a_1, \end{aligned}$$

and

$$m_{ij} = 0 \quad \text{for any} \quad i = 5, 6, \cdots, 26. \tag{8}$$

The proof of Lemma 1 is essentially the same as that of Lemma 1 of I.

For the rest of the paper, we assume the restriction $a_1 \leq a_3$ on the coordinates of N_h in (6), which eliminates from our consideration a class of IR's of F_4. In the last section, we discuss how to treat this class using different restrictions on the coordinates of the highest weight.

According to (c) of Lemma 1, we can use A_{25}-patterns for construction of the basis in $R(\Psi)$ that are of the form

$$(g) = \begin{pmatrix} a_2+a_3+a_4 & a_2+a_3 & a_1+a_2 & a_1 & 0 & \cdots & 0 \\ & m_{1,25} & m_{2,25} & m_{3,25} & m_{4,25} & 0 & \cdots & 0 \\ & & m_{1,24} & m_{2,24} & m_{3,24} & m_{4,24} & 0 & \cdots & 0 \\ & & & & \cdot & & & & \\ & & & & \cdot & & & & \\ & & & & \cdot & & & & \\ & & & & m_{1,1} & & & & \end{pmatrix}. \tag{9}$$

The A_{25}-weight of (9) is given by (I5). Its F_4-weight N follows from (I5), (4), and (8):

$$\begin{aligned} N = {} & \tau_1\Big(-\sum_{k=1}^{3} m_{k,3} + \sum_{k=1}^{4}(2m_{k,5} - 2m_{k,6} + 2m_{k,7} - m_{k,9} - m_{k,17} + 2m_{k,19} - 2m_{k,20} + 2m_{k,21} - m_{k,23})\Big) \\ & + \tau_2\Big(-\sum_{k=1}^{2} m_{k,2} + 2\sum_{k=1}^{3} m_{k,3} + \sum_{k=1}^{4}(-m_{k,4} - m_{k,7} + 2m_{k,9} - m_{k,11} \\ & - m_{k,15} + 2m_{k,17} - m_{k,19} - m_{k,22} + 2m_{k,23} - m_{k,24})\Big) \\ & + \tau_3\Big(-m_{11} + 2\sum_{k=1}^{2} m_{k,2} - 2\sum_{k=1}^{3} m_{k,3} + \sum_{k=1}^{4}(2m_{k,4} - 2m_{k,5} + m_{k,6} + m_{k,7} - m_{k,8} - m_{k,9} - m_{k,10} \\ & + 3m_{k,11} - m_{k,12} - m_{k,14} + 3m_{k,15} - m_{k,16} - m_{k,17} - m_{k,18} \\ & + m_{k,19} + m_{k,20} - 2m_{k,21} + 2m_{k,22} - 2m_{k,23} + 2m_{k,24} - m_{k,25})\Big) \\ & + \tau_4\Big(2m_{11} - \sum_{k=1}^{2} m_{k,2} + \sum_{k=1}^{4}(-m_{k,4} + m_{k,5} + m_{k,6} - 2m_{k,7} + 2m_{k,8} - 2m_{k,9} + 2m_{k,10} \\ & - 3m_{k,11} + 2m_{k,12} + 2m_{k,14} - 3m_{k,15} + 2m_{k,16} - 2m_{k,17} + 2m_{k,18} - 2m_{k,19} \\ & + m_{k,20} + m_{k,21} - m_{k,22} - m_{k,24} + 2m_{k,25} - 2a_1 - 3a_2 - 2a_3 - a_4)\Big). \end{aligned} \tag{10}$$

III. LABELING STATES AND CONSTRUCTION OF REPRESENTATIONS

Let us write the generators of the subalgebra F_4 of A_{25} as linear combinations of A_{25}-generators $I_{i,j}$ and $I_{i,i} - I_{j,j}$, $i \neq j$, of Gel'fand and Tsetlin.[2] The lowering generators[6] corresponding to negative simple roots of F_4 are

$$\begin{aligned} E_{-\beta_1} &= I_{6,4} + I_{8,5} + I_{9,7} + I_{20,18} + I_{22,19} + I_{232,1}, \\ E_{-\beta_2} &= I_{4,3} + I_{10,8} + I_{11,9} + I_{18,16} + I_{19,17} + I_{24,23}, \\ E_{-\beta_3} &= 2^{-\frac{1}{2}}[I_{3,2} + I_{5,4} + I_{8,6} + I_{12,10} + a(I_{13,11} + I_{16,13}) + b(I_{14,11} + I_{16,14}) + I_{17,15} + I_{21,19} + I_{23,22} + I_{25,24}], \\ E_{-\beta_4} &= 2^{-\frac{1}{2}}[I_{2,1} + I_{7,5} + I_{9,8} + I_{11,10} + b(I_{13,12} + I_{15,13}) + a(I_{14,12} + I_{15,14}) + I_{17,16} + I_{19,18} + I_{22,20} + I_{26,25}], \end{aligned} \tag{11}$$

where $a = (1 + \sqrt{\frac{3}{4}})^{\frac{1}{2}}$ and $b = (1 - \sqrt{\frac{3}{4}})^{\frac{1}{2}}$, as can be verified, for instance, by comparing nonzero commutators of $E_{-\beta_i}$ with the system of roots of F_4. Indeed, the lowering generators corresponding to nonsimple roots of F_4 are proportional to commutators of $E_{-\beta_i}$, i.e.,

$$[E_{-\beta_i}, E_{-\beta_j}] \sim E_{-\beta_i-\beta_j}. \tag{12}$$

Thus one gets the generators

$$\begin{aligned} E_{-\beta_1-\beta_2} &= I_{6,3} - I_{10,5} - I_{11,7} + I_{20,16} + I_{22,17} - I_{24,21}, \\ E_{-\beta_2-\beta_3} &= 2^{-\frac{1}{2}}[I_{4,2} - I_{5,3} + I_{10,6} - I_{12,8} - a(I_{13,9} - I_{18,13}) \\ &\quad - b(I_{14,9} - I_{18,14}) + I_{19,15} - I_{21,17} + I_{24,22} - I_{25,23}], \\ E_{-\beta_3-\beta_4} &= 2^{-\frac{1}{2}}[I_{3,1} - I_{7,4} - I_{9,6} + (a-b)(I_{13,10} - I_{17,13} - I_{14,10} + I_{17,14}) \\ &\quad - I_{15,11} + I_{16,12} + I_{21,18} + I_{23,20} - I_{26,24}], \end{aligned}$$

$$
\begin{aligned}
E_{-\beta_1-\beta_2-\beta_3} &= 2^{-\frac{1}{2}}[I_{6,2} - I_{8,3} - I_{10,4} + I_{12,5} + a(I_{13,7} + I_{20,13}) \\
&\quad + b(I_{14,7} + I_{20,14}) + I_{22,15} - I_{23,17} - I_{24,19} + I_{25,21}], \\
E_{-\beta_2-\beta_3-\beta_4} &= 2^{-\frac{1}{2}}[I_{4,1} + I_{7,3} - I_{11,6} + I_{15,9} - (a-b)(I_{13,8} + I_{19,13} \\
&\quad - I_{14,8} - I_{19,14}) + I_{18,12} - I_{21,16} + I_{24,20} + I_{26,23}], \\
E_{-\beta_2-2\beta_3} &= -I_{5,2} - I_{12,6} + I_{16,9} + I_{18,11} - I_{21,15} - I_{25,22}, \\
E_{-\beta_1-\beta_2-\beta_3-\beta_4} &= 2^{-\frac{1}{2}}[I_{6,1} + I_{11,4} + (a-b)(I_{13,5} - I_{22,13} - I_{14,5} + I_{22,14}) \\
&\quad + I_{9,3} - I_{15,7} + I_{20,12} - I_{24,18} - I_{23,16} - I_{26,21}], \\
E_{-\beta_1-\beta_2-2\beta_3} &= -I_{8,2} + I_{12,4} + I_{20,11} - I_{16,7} - I_{23,15} + I_{25,16}, \\
E_{-\beta_2-2\beta_3-\beta_4} &= 2^{-\frac{1}{2}}[-I_{5,1} + I_{7,2} + I_{18,10} + I_{16,8} + b(I_{13,6} - I_{21,13}) \\
&\quad + a(I_{14,6} - I_{21,14}) - I_{17,9} - I_{19,11} - I_{25,20} + I_{26,22}], \\
E_{-\beta_1-\beta_2-2\beta_3-\beta_4} &= 2^{-\frac{1}{2}}[-b(I_{13,4} + I_{23,13}) - a(I_{14,4} + I_{23,14}) - I_{8,1} - I_{16,5} \\
&\quad + I_{9,2} + I_{17,7} + I_{20,10} + I_{25,18} - I_{22,11} - I_{26,19}], \\
E_{-\beta_1-2\beta_2-2\beta_3} &= I_{10,2} + I_{12,3} + I_{20,9} + I_{18,7} + I_{25,17} + I_{24,15}, \\
E_{-\beta_2-2\beta_3-2\beta_4} &= I_{7,1} - I_{17,8} - I_{19,10} - I_{21,12} - I_{15,6} + I_{26,20}, \\
E_{-\beta_1-\beta_2-2\beta_3-2\beta_4} &= I_{9,1} + I_{15,4} + I_{17,5} - I_{22,10} - I_{23,12} - I_{26,18}, \\
E_{-\beta_1-2\beta_2-2\beta_3-\beta_4} &= 2^{-\frac{1}{2}}[I_{10,1} + I_{18,5} + I_{20,8} - I_{11,2} - b(I_{13,3} - I_{24,13}) \\
&\quad - a(I_{14,3} - I_{24,14}) + I_{25,16} - I_{19,7} - I_{22,9} - I_{26,17}], \\
E_{-\beta_1-2\beta_2-2\beta_3-2\beta_4} &= I_{15,3} - I_{22,8} - I_{11,1} - I_{26,16} - I_{19,5} + I_{24,12}, \\
E_{-\beta_1-2\beta_2-3\beta_3-\beta_4} &= 2^{-\frac{1}{2}}[(a-b)(I_{13,2} + I_{25,13} - I_{14,2} - I_{25,14}) + I_{18,4} + I_{20,6} \\
&\quad - I_{12,1} + I_{24,11} - I_{26,15} + I_{21,7} + I_{23,9} + I_{16,3}], \\
E_{-\beta_1-2\beta_2-3\beta_3-2\beta_4} &= 2^{-\frac{1}{2}}[-I_{17,3} + I_{15,2} - I_{19,4} - I_{22,16} + I_{24,10} + a(I_{13,1} - I_{26,13}) \\
&\quad + b(I_{14,1} - I_{26,14}) + I_{21,5} + I_{23,8} - I_{25,12}], \\
E_{-\beta_1-2\beta_2-4\beta_3-2\beta_4} &= -I_{17,2} + I_{21,4} - I_{26,11} - I_{16,1} + I_{23,6} - I_{25,10}, \\
E_{-\beta_1-3\beta_2-4\beta_3-2\beta_4} &= I_{19,2} + I_{18,1} + I_{21,3} - I_{24,6} - I_{25,8} - I_{26,9}, \\
E_{-2\beta_1-3\beta_2-4\beta_3-2\beta_4} &= -(I_{20,1} + I_{22,2} + I_{23,3} + I_{24,4} + I_{25,5} + I_{26,7}).
\end{aligned} \tag{13}
$$

The raising generators E_γ, where γ is one of the positive roots of F_4, are obtained from $E_{-\gamma}$ by permutation of the indices of each $I_{i,j}$.

The generators H_γ, elements of the Cartan subalgebra of F_4, are found as[6]

$$H_\gamma = [E_\gamma, E_{-\gamma}]. \tag{14}$$

In particular,

$$
\begin{aligned}
H_{\beta_1} &= I_{4,4} + I_{5,5} - I_{6,6} + I_{7,7} - I_{8,8} - I_{9,9} + I_{18,18} + I_{19,19} - I_{20,20} \\
&\quad + I_{21,21} - I_{22,22} - I_{23,23}, \\
H_{\beta_2} &= I_{3,3} - I_{4,4} + I_{8,8} + I_{9,9} - I_{10,10} - I_{11,11} + I_{16,16} + I_{17,17} \\
&\quad - I_{18,18} - I_{19,19} + I_{23,23} - I_{24,24}, \\
H_{\beta_3} &= \tfrac{1}{2}(I_{2,2} - I_{3,3} + I_{4,4} - I_{5,5} + I_{6,6} - I_{8,8} + I_{10,10} + 2I_{11,11} - I_{12,12} + I_{15,15} \\
&\quad - 2I_{16,16} - I_{17,17} + I_{19,19} - I_{21,21} + I_{22,22} - I_{23,23} + I_{24,24} - I_{25,25}), \\
H_{\beta_4} &= \tfrac{1}{2}(I_{1,1} - I_{2,2} + I_{5,5} - I_{7,7} + I_{8,8} - I_{9,9} + I_{10,10} - I_{11,11} + 2I_{12,12} \\
&\quad - 2I_{15,15} + I_{16,16} - I_{17,17} + I_{18,18} - I_{19,19} + I_{20,20} - I_{22,22} + I_{25,25} - I_{26,26}).
\end{aligned} \tag{15}
$$

All representations of A_{25}-generators are known,[2] in particular, the representation $\Phi(A_{25})$ of Lemma 1. Because of the one-to-one correspondence between an element of an algebra and its representation, one is free to consider $I_{i,j}$ and $I_{i,i} - I_{j,j}$ as A_{25}-generators in the representation $\Phi(A_{25})$. Consequently, (11), (13), and (15) are generators of $\Phi(F_4)$.

Our solution of the state labeling problem for the class (6) of IR's of F_4 is given in

Lemma 2: Let $\Psi(F_4)$ be the IR of F_4 with the highest weight (6). Then a complete basis in the representation space $R(\Psi)$ consists of the vector

$$(g_h) = \begin{pmatrix} a_2+a_3+a_4 & a_2+a_3 & a_1+a_2 & a_1 & 0 & \cdots & 0 \\ & a_2+a_3+a_4 & a_2+a_3 & a_1+a_2 & a_1 & 0 & \cdots & 0 \\ & & \vdots & & & & & \\ & & a_2+a_3+a_4 & & a_2+a_3 & & & \\ & & & a_2+a_3+a_4 & & & & \end{pmatrix} \tag{16}$$

and all linearly independent vectors

$$E_{-\beta_{i(k)}} E_{-\beta_{i(k-1)}} \cdots E_{-\beta_{i(2)}} E_{-\beta_{i(1)}} (g_h), \tag{17}$$

where $1 \leq k \leq 22a_1 + 42a_2 + 30a_3 + 16a_4$ and $i(k) = 1, 2, 3$, or 4.

The proof of Lemma 2 is essentially the same as that of Lemma 2 of I.

Finally, the matrix elements of F_4-generators in the IR $\Psi(F_4)$ are obtained when (11), (13), and (15) are applied to the basis constructed in Lemma 2.

IV. EXAMPLE

As an example, let us calculate a matrix element

$$W_{fi} = \langle (g^f) | E_{-\beta_1-\beta_2} | (g^i) \rangle \tag{18}$$

between the normalized states

$$|(g^i)\rangle = (N_i)^{-1}(g_h) \quad \text{and} \quad |(g^f)\rangle = (N_f)^{-1} E_{-\beta_1} E_{-\beta_2} (g_h) \tag{19}$$

of the 1274-dimensional representation Ψ of F_4 whose highest weight is $N_h = \tau_2$.

According to Lemma 1, the highest weight of $\Phi(A_{25})$, whose patterns we have to use in our example, is $M_h = \nu_3$. Consequently, the pattern (g_h) of Lemma 2 is

$$(g_h) = \begin{pmatrix} 1 & 1 & 1 & 0 & \cdots & 0 \\ & & \vdots & & & \\ & 1 & 1 & 1 & 0 & \\ & & 1 & 1 & 1 & \\ & & & 1 & 1 & \\ & & & & 1 & \end{pmatrix}. \tag{20}$$

Let us first verify that the vectors (19) do belong to the representation space $R(\Psi)$. According to Lemma 2, it is obvious for $|(g^i)\rangle$, and one has to show it only for $|(g^f)\rangle$. The straightforward proof would be to substitute (11) and (20) into (19) and use the standard rules[2] about how the operators $I_{i,j}$ act on Gel'fand patterns. However, it is shorter to verify only that each successive application of a lowering operator $E_{-\beta_i}$ in (19) gives a vector whose F_4-weight belongs to the weight system of $\Psi(F_4)$. Thus, in our example, one has to check that $N_h - \beta_2$ and $N_h - \beta_2 - \beta_1$ are weights of Ψ. This is easily done using (1) and the algorithm for calculation of weights.[5]

Now we can start the calculation of (18). Again it is possible to substitute for $E_{-\beta_1-\beta_2}$ the corresponding linear transformation (13) of operators $I_{i,j}$ and let them act on (g_h). However, it is much shorter if we first make several simplifications. Using (11) and the commutation relations[1,2] of $I_{i,j}$, one finds

$$E_{-\beta_1-\beta_2} = [E_{-\beta_1}, E_{-\beta_2}].$$

Hence,

$$\begin{aligned} W_{fi} &= (N_i N_f)^{-1} \langle E_{-\beta_1} E_{-\beta_2} (g_h) | \, E_{-\beta_1} E_{-\beta_2} \\ &\qquad - E_{-\beta_2} E_{-\beta_1} | (g_h) \rangle \\ &= (N_i N_f)^{-1} \langle (g_h) | \, E_{\beta_2} E_{\beta_1} (E_{-\beta_1} E_{-\beta_2} \\ &\qquad - E_{-\beta_2} E_{\beta_1}) | (g_h) \rangle. \end{aligned} \tag{21}$$

Using the commutation relations (14) and[6]

$$[H_{\beta_i}, E_{\beta_j}] = (\beta_i, \beta_j) E_{\beta_j}, \quad H_{\beta_i} | (g_h) \rangle = (\beta_i, N_h) | (g_h) \rangle \tag{22}$$

together with

$$\cdots E_{\beta_i} | (g_h) \rangle = \langle (g_h) | \, E_{-\beta_j} \cdots = 0, \tag{23}$$

one has

$$\begin{aligned} W_{fi} &= \frac{1}{N_i N_f} \langle (g_h) | \, -(\beta_1, \beta_2) H_{\beta_2} | (g_h) \rangle \\ &= -\frac{(\beta_1, \beta_2)}{N_i N_f} (\beta_2, N_h) \langle g_h \mid g_h \rangle \\ &= \frac{1}{N_i N_f} (\beta_2, \tau_2) \langle (g_h) \mid (g_h) \rangle = \frac{\langle (g_h) \mid (g_h) \rangle}{N_i N_f}. \end{aligned} \tag{24}$$

Here we have used the property $(\beta_1, \beta_2) = -1$ of simple roots of F_4 and (11). Similarly one can calculate also the normalization coefficients N_i and N_f. Obviously the normalization coefficient N_i is given by the normalization of A_{25}-Gel'fand-patterns, i.e.,

$$N_i = [\langle (g_h) \mid (g_h)\rangle]^{\frac{1}{2}}, \tag{25}$$

and

$$\begin{aligned} N_f &= [\langle E_{-\beta_1}E_{-\beta_2}(g_h) \mid E_{-\beta_1}E_{-\beta_2}(g_h)\rangle]^{\frac{1}{2}} \\ &= [\langle (g_h)| E_{\beta_2}E_{\beta_1}E_{-\beta_1}E_{-\beta_2} |(g_h)\rangle]^{\frac{1}{2}} \\ &= [\langle (g_h)| H_{\beta_1}H_{\beta_2} - (\beta_1, \beta_2)H_{\beta_2} |(g_h)\rangle]^{\frac{1}{2}} \\ &= \{\langle (g_h) \mid (g_h)\rangle(\beta_2, \tau_2)[1 + (\beta_1, \tau_2)]\rangle\}^{\frac{1}{2}} \\ &= [\langle (g_h) \mid (g_h)\rangle]^{\frac{1}{2}}. \end{aligned} \tag{26}$$

Finally, using (25) and (26) in (24), we have $W_{fi} = 1$.

V. REMARKS

The inequality $a_1 \leq a_3$ of (6) restricts the class of F_4-representation to which the complete labeling of Lemma 2 can be applied. Indeed, were $a_1 > a_3$, then the coordinate $(a_3 - a_1)$ at v_2 of the highest weight M_h of (7) would be negative. However, the basis $\{v_i\}$ is constructed in such a way that all coordinates of any highest weight of any representation must be non-negative, as can be seen, for instance, by comparing (I4), (I5′), and (7).

It is worthwhile to point out that some restriction similar to (6) seems fundamental and therefore cannot be avoided. However, it can be modified either by choosing a "bigger" representation of A_{25} than that of (7) or by embedding F_4 into another algebra of rank higher than 25: thus, for instance, replacing (7) by

$$M_h = (a_4 - \tfrac{1}{2}a_1)v_1 + a_3v_2 + a_2v_3 + \tfrac{1}{2}a_1v_5. \tag{27}$$

Lemma 1 holds provided that N_h of (6) is subject of the condition

$$a_1 \text{ even} \quad \text{and} \quad a_4 \geq \tfrac{1}{2}a_1, \tag{28}$$

(b) of the Lemma is omitted, and (c) is modified. In fact, even the most general form (4) of M_h in (7) would not allow us to consider all IR's of $\Psi(F_4)$ in the present embedding.

For many applications it is useful as a check to know independently the normalization of the representations. The matrices $E_{\pm\gamma_i}$ and H_{γ_i}, where γ_i is a root of F_4, are the generators of the (reducible) representation $\Phi(F_4)$. They are normalized as

$$\begin{aligned} (E_{\pm\gamma_i})^2 &= \operatorname{Tr} E_{\gamma_i}E_{-\gamma_i} = j_{F_4} \operatorname{Tr} I_{pq}I_{qp} = j_{F_4}l_{\Phi(A_{25})} \\ &= l_{\Phi(F_4)} = l_{\Psi(F_4)} + l_{\Psi_1(F_4)} + \cdots, \\ (H_{\gamma_i})^2 &= \operatorname{Tr} H_{\gamma_i}H_{\gamma_i} = (\gamma_i, \gamma_i)j_{F_4}l_{\Phi(A_{25})} \\ &= (\gamma_i, \gamma_i)(l_{\Psi(F_4)} + l_{\Psi_1(F_4)} + \cdots). \end{aligned} \tag{29}$$

Here $\Psi(F_4), \Psi_1(F_4), \cdots$ are the IR's in $\Phi(F_4)$; $j_{F_4} = 6$ is the index of the subalgebra F_4 in[7,8] A_{25}; $l_{\Phi(G)}$ is the index of the representation Φ of the algebra G.[7,9] Hence, in the IR $\Psi(F_4)$, the generators are normalized as

$$\begin{aligned} (E_{\pm\gamma_i})^2 &= \operatorname{Tr} E_{\gamma_i}E_{-\gamma_i} = l_{\Psi(F_4)}, \\ (H_{\gamma_i})^2 &= \operatorname{Tr} H_{\gamma_i}H_{\gamma_i} = (\gamma_i, \gamma_i)l_{\Psi(F_4)}. \end{aligned} \tag{30}$$

The value of $l_{\Psi(F_4)}$ is calculated from[9]

$$\begin{aligned} l_\Psi = [d(\Psi)/52](&2a_1^2 + 6a_2^2 + 3a_3^2 + a_4^2 \\ &+ 6a_1a_2 + 4a_1a_3 + 2a_1a_4 + 8a_2a_3 + 4a_2a_4 \\ &+ 3a_3a_4 + 16a_1 + 30a_2 + 21a_3 + 11a_4), \end{aligned} \tag{31}$$

where a_i are the coordinates (6) of the highest weight N_h of $\Psi(F_4)$ and $d(\Psi)$ is the dimension[5] of Ψ.

ACKNOWLEDGMENT

I am grateful for the hospitality extended to me at the Aspen Center for Physics, where most of this work was done.

[1] J. Patera, J. Math. Phys. **11**, 3027 (1970).
[2] I. M. Gel'fand and M. L. Tsetlin, Dokl. Akad. Nauk SSSR **71**, 825 (1950); also I. M. Gel'fand, R. A. Minlos, and Z. Ya. Shapiro, *Representations of the Rotation and Lorentz Groups and Their Applications* (Pergamon, New York, 1963), Suppl. II. The results were rederived and signs of final formulas corrected by G. E. Baird and L. C. Biedenharn, J. Math. Phys. **4**, 1449 (1963).
[3] The same difficulty arises when considering the algebras of types E_6, E_7, and E_8.
[4] A. Navon and J. Patera, J. Math. Phys. **8**, 489 (1967).
[5] E. B. Dynkin, Am. Math. Soc. Transl., Ser. 2, **6**, 245 (1957).
[6] For instance, Ref. 5, p. 326.
[7] E. B. Dynkin, Am. Math. Soc. Transl., Ser. 2, **6**, 111 (1957), Tables 16 and 25.
[8] J. Patera, Nuovo Cimento **46**, 637 (1966).
[9] J. Patera, Nuovo Cimento **58A**, 402 (1968).

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 12, NUMBER 3 MARCH 1971

Group Theory of Superfluidity

A. I. SOLOMON
Dublin Institute for Advanced Studies, Dublin, Ireland

(Received 1 June 1970)

The method of using a dynamical group to generate the energy spectrum of a given Hamiltonian is applied to the case of a superfluid Bose system. Here the relevant group is found to be $SU(1,1)$ [or $\prod_k \otimes SU(1,1)_k$ for a multilevel system]. The energy eigenvalues and eigenfunctions are obtained by means of the group.

1. INTRODUCTION

We describe a method for treating a system of superfluid bosons which is based on similar treatments of the hydrogen atom.[1] This method exploits the use of groups which are not themselves symmetry groups of the Hamiltonian, but whose algebras are the so-called spectrum generating algebras.[2] A knowledge of the relevant group of the problem enables one to write down the energy eigenvalues and evaluate the eigenstates, as well as suggesting what form additional correction terms to the Hamiltonian should take. We illustrate the method by first applying it to a simple superfluid system consisting only of a ground level and a degenerate excited level (the Foldy model[3]); we shall see that the relevant group here is the noncompact pseudo-unitary group $SU(1,1)$. This then enables us to treat a many-level superfluid system, since in the usual approximation the reduced Hamiltonian is so chosen that the spectrum generating group is simply a direct product $\prod_k \otimes SU(1,1)_k$, which introduces essentially no additional complication. We exhibit the energy spectrum and eigenstates for this case.

2. SIMPLE SUPERFLUID MODEL

Let us consider a system of N weakly interacting bosons described by the following Hamiltonian:

$$H = \sum_k \epsilon_k a_k^\dagger a_k + \tfrac{1}{2}\sum_{k,p,q} V_k a_{p+k}^\dagger a_{q-k}^\dagger a_p a_q, \quad \epsilon_k = k^2/2m. \tag{2.1}$$

If we limit ourselves to the case where there are only three states in the system, so that $p, q, k, \cdots$ take on the values $(-1, 0, +1)$ only and such that $\epsilon_{\pm 1} = \epsilon$, $V_{\pm 1} = V$, and $\epsilon_0 = 0 = V_0$ (the Foldy model[3]), then the Hamiltonian becomes

$$H = \epsilon(a_+^\dagger a_+ + a_-^\dagger a_-) + V[a_0^\dagger a_0(a_+^\dagger a_+ + a_-^\dagger a_-) + a_0^{\dagger 2} a_+ a_- + a_0^2 a_+^\dagger a_-^\dagger] \tag{2.2}$$

(where we have written $a_\pm$ for $a_{\pm 1}$). For $V = 0$, the ground state would consist of all N particles in the zero state; let us assume for our weakly interacting system that the zero state is macroscopically occupied, so that we may treat the operators a_0 and $a_0^\dagger$ as if they were equal to the c number $N_0^{\frac{1}{2}}$, where $N_0 = \langle a_0^\dagger a_0 \rangle$ (ground state). This is the physical assumption which gives rise to the superfluid character of the model. Thus the reduced Hamiltonian is

$$H_{\text{red}} = (\epsilon + N_0 V)(a_+^\dagger a_+ + a_-^\dagger a_-) + N_0 V(a_+^\dagger a_-^\dagger + a_+ a_-). \tag{2.3}$$

Now denote the operators occurring in H_{red} by

$$J_1 = -\tfrac{1}{2}(a_+^\dagger a_-^\dagger + a_+ a_-)$$

and

$$J_3 = \tfrac{1}{2}(a_+^\dagger a_+ + a_-^\dagger a_- + 1).$$

We then find that the algebra generated by commutation of these operators closes on the introduction of only one additional operator

$$J_2 = \tfrac{1}{2}i(a_+^\dagger a_-^\dagger - a_+ a_-).$$

These Hermitian operators have the commutation rules

$$[J_1, J_2] = -iJ_3, \quad [J_2, J_3] = iJ_1, \quad [J_3, J_1] = iJ_2. \tag{2.4}$$

If the first commutator had a plus sign, this would be the well-known angular momentum algebra of

$$SO(3) \sim SU(2);$$

with the above signs the algebra is that of the noncompact pseudo-unitary group $SO(2,1) \sim SU(1,1)$.

In terms of the $SU(1,1)$ generators (2.4), the reduced Hamiltonian (2.3) becomes

$$H_{\text{red}} = 2N_0 V(\mu J_3 - J_1 - \tfrac{1}{2}\mu), \quad \mu = 1 + \epsilon/N_0 V. \tag{2.5}$$

The problem now consists in solving the Schrödinger equation

$$H|\psi_n\rangle = E_n|\psi_n\rangle. \tag{2.6}$$

This may be considerably simplified if we note that one of the operators J_1 or J_3 may be rotated away,

for consider a rotation by θ about the J_2 axis,

$$R(\theta) \equiv e^{-iJ_2\theta},$$
$$R(\theta)J_1R^{-1}(\theta) = \cosh\theta J_1 + \sinh\theta J_3,$$
$$R(\theta)J_3R^{-1}(\theta) = \cosh\theta J_3 + \sinh\theta J_1,$$

so that

$$R(\theta)HR^{-1}(\theta) = 2N_0V[J_3(\mu\cosh\theta - \sinh\theta) - J_1(\cosh\theta - \mu\sinh\theta) - \tfrac{1}{2}\mu].$$

Since $|\tanh\theta| < 1$ for all θ, we may rotate either J_1 or J_3 away depending on the sign of the potential V; thus

(i) $V < 0$; attractive potential, $\mu < 1$: choose $\theta = \tanh^{-1}\mu$,

$$RHR^{-1} = -2N_0V(J_1\,\mathrm{sech}\,\theta + \tfrac{1}{2}\mu);$$

(ii) $V > 0$; repulsive potential, $\mu > 1$: choose $\theta = \coth^{-1}\mu$

$$RHR^{-1} = 2N_0V(J_3\,\mathrm{csch}\,\theta - \tfrac{1}{2}\mu).$$

Since

$$[R(\theta)HR^{-1}(\theta)]R(\theta)|\psi_n\rangle = E_nR(\theta)|\psi_n\rangle, \quad (2.7)$$

the energy spectrum is immediately obtained as that of the rotated Hamiltonian and the eigenstates as the rotated eigenstates of the diagonalized generator. Thus, in case (i) we obtain a continuous spectrum, since that of the noncompact generator J_1 is continuous. The second case is of more physical interest, as the spectrum of the compact generator J_3 is discrete and integer spaced. The Casimir operator may be written

$$C = -J_1^2 - J_2^2 + J_3^2 = \tfrac{1}{4}(\Delta^2 - 1)$$

in terms of the invariant of the Hamiltonian

$$\Delta = a_+^\dagger a_+ - a_-^\dagger a_-.$$

Since our spectrum must be bounded below, the only allowed representation is $\mathfrak{D}^+(j)$ [see Appendix, cases (A13b) and (A14b)]. For a unitary representation,

$$j = -\tfrac{1}{2} - \tfrac{1}{2}|\Delta| = -\sigma, \quad \text{for a given } \Delta.$$

Thus the spectrum is given by

$$J_3|n\rangle = (n + \sigma)|n\rangle \quad (2.8)$$

or

$$E_n = (2n + 1 + |\Delta|)E - N_0V - \epsilon,$$

where

$$E = (2\epsilon N_0V + \epsilon^2)^{\frac{1}{2}}, \quad n = 0, 1, 2, \cdots. \quad (2.9)$$

The energy eigenstates are obtained from Eqs. (2.7) and (2.8) as

$$|\psi_n\rangle = R^{-1}(\theta)|n\rangle$$

or

$$|\psi_n\rangle = \sum_m |m\rangle\langle m|\,R^{-1}(\theta)\,|n\rangle = \sum_m S^j_{mn}(g)\,|m\rangle$$

in the notation of the Appendix. The group element corresponding to the finite rotation $R^{-1}(\theta)$ is given by $g(\alpha, \beta)$, where $\alpha = \cosh\frac{1}{2}\theta$, $\beta = \sinh\frac{1}{2}\theta$, by use of the representation (A5). Then $S^j_{mn}(g)$ is given by expression (A15).

As an example consider the case $|\Delta| = 0$ (equal numbers of "+" and "−" particles). The ground state $|\psi_0\rangle = \sum_m S^{-\frac{1}{2}}_{m0}(g)\,|m\rangle$ may be calculated from (A15) by using the properties of gamma functions and noting that only one term contributes to the sum, and one finds

$$|\psi_0\rangle = \sum_m (-)^m\,\mathrm{sech}\,\tfrac{1}{2}\theta(\tanh\tfrac{1}{2}\theta)^m\,|m\rangle. \quad (2.10)$$

In terms of the free-particle vacuum state $|0\rangle$ obeying $a_+|0\rangle = a_-|0\rangle = 0$, $|m\rangle = (m!)^{-1}(a_+^\dagger a_-^\dagger)^m|0\rangle$, this simplifies to

$$|\psi_0\rangle = (1 - t^2)^{\frac{1}{2}}\exp(-ta_+^\dagger a_-^\dagger)\,|0\rangle, \quad (2.11)$$

where

$$t = \tanh\tfrac{1}{2}\theta = -E_0/N_0V.$$

The results (2.9) and (2.11) are those obtained by diagonalization[3] via the Bogoliubov transformation.[4]

3. SUPERFLUID HELIUM

In view of the preceding section it is now a simple matter to treat the Hamiltonian (2.1). Using, as before, the Bogoliubov approximation $a_0 \sim a_0^\dagger \sim N_0^{\frac{1}{2}}$, we may write (2.1) as

$$H = \tfrac{1}{2}N_0^2V_0 + \sum(\epsilon_k + N_0V_k + N_0V_0)a_k^\dagger a_k + \tfrac{1}{2}N_0\sum V_k(a_k^\dagger a_{-k}^\dagger + a_ka_{-k}),$$

where the summation is over k and does not include $k = 0$, and we neglect terms of higher order in a_k. To the same order we may use $N = N_0 + \sum a_k^\dagger a_k$, in the sense of the expectation values, to rewrite H as

$$H = \tfrac{1}{2}N^2V_0 + \sum(\epsilon_k + NV_k)a_k^\dagger a_k + \tfrac{1}{2}N\sum V_k(a_k^\dagger a_{-k}^\dagger + a_ka_{-k}). \quad (3.1)$$

Writing

$$J_1^{(k)} = -\tfrac{1}{2}(a_k^\dagger a_{-k}^\dagger + a_ka_{-k}),$$
$$J_2^{(k)} = \tfrac{1}{2}i(a_k^\dagger a_{-k}^\dagger - a_ka_{-k}), \quad (3.2)$$
$$J_3^{(k)} = \tfrac{1}{2}(a_k^\dagger a_k + a_{-k}^\dagger a_{-k} + 1),$$

we see that $\{J_1^{(k)}, J_2^{(k)}, J_k^{(3)}\}$ generate the algebra of the group $SU(1, 1)_{(k)}$, and the Hamiltonian (3.1) may be written in terms of the direct sum of generators $J_i^{(k)}$ of the product group $\prod_k \otimes SU(1, 1)_{(k)}$; thus

$$H = \sum_k \oplus NV_k(-J_1^{(k)} + \mu_kJ_2^{(k)} - \tfrac{1}{2}\mu_k) + \tfrac{1}{2}N^2V_0, \quad \mu_k = 1 + \epsilon_k/NV_k. \quad (3.3)$$

As before, the case $V_k < 0$, $\mu_k > 1$ gives the continuous spectrum, while for $V_k > 0$, or $\mu_k < 1$, the Hamiltonian may be simplified by the rotation

$$R = \prod_k \otimes R(\theta_k), \quad R(\theta_k) = \exp(-iJ_2^{(k)}\theta_k),$$
$$\theta_k = \coth^{-1}\mu_k, \tag{3.4}$$

so that

$$RHR^{-1} = \sum_k \oplus (\operatorname{csch}\theta_k J_3^{(k)} - \tfrac{1}{2}\mu_k)NV_k + \tfrac{1}{2}N^2V_0. \tag{3.5}$$

The Casimir invariants are

$$C_k = -(J_1^{(k)})^2 - (J_2^{(k)})^2 + (J_3^{(k)})^2 = \tfrac{1}{4}(\Delta_k^2 - 1),$$

where the constants of the motion $\Delta_k = a_k^\dagger a_k - a_{-k}^\dagger a_{-k}$ are the differences between numbers of particles in opposite momentum states. As the spectrum must be bounded below, the only allowed representation is

$$\prod_k \otimes \mathfrak{D}^+(j_k), \quad j_k = -\tfrac{1}{2} - \tfrac{1}{2}|\Delta_k| = -\sigma_k.$$

Writing the basis in terms of eigenstates of $J_3^{(k)}$, where $J_3^{(k)}|n_k\rangle = (n_k + \sigma_k)|n_k\rangle$, so that

$$|n_1, n_2, \cdots, n_k, \cdots\rangle \equiv |n_1\rangle|n_2\rangle\cdots|n_k\rangle\cdots,$$
$$n_k = 0, 1, 2, \cdots, \tag{3.6}$$

we see that the energy eigenvalues are

$$E(n_1, n_2, \cdots, n_k, \cdots) = \sum(n_k + \tfrac{1}{2} + \tfrac{1}{2}|\Delta_k|)E_k + \text{const},$$
$$\text{where} \quad E_k = (2\epsilon_k NV_k + \epsilon_k^2)^{\frac{1}{2}}. \tag{3.7}$$

The energy eigenfunctions may also be written down explicitly as before; thus

$$H|\psi(n_1, n_2, \cdots, n_k, \cdots)\rangle = E(n_1, n_2, \cdots, n_k, \cdots) \times |\psi(n_1, n_2, \cdots, n_k, \cdots)\rangle,$$

where

$$|\psi(n_1, n_2, \cdots, n_k, \cdots)\rangle = \prod_k \otimes \sum_{m_k} S^{j_k}_{m_k n_k}(g_k)|m_k\rangle,$$

where

$$g_k = g(\alpha_k, \beta_k) = \begin{bmatrix} \alpha_k & \beta_k \\ \bar{\beta}_k & \bar{\alpha}_k \end{bmatrix},$$
$$\alpha_k = \cosh\tfrac{1}{2}\theta_k \quad \beta_k = \sinh\tfrac{1}{2}\theta_k,$$

Again the ground state may be calculated explicitly to yield

$$|\psi(0, 0, \cdots)\rangle = \left(\prod_k (1 - t_k^2)^{\frac{1}{2}}\right)\exp\left(\sum_k (-t_k a_k^\dagger a_{-k}^\dagger)|0\rangle\right)$$

for the case $\Delta_k = 0$, $t_k \equiv \tanh\tfrac{1}{2}\theta_k$.

4. SUMMARY

The method we have utilized in this paper, the application of groups which are not symmetry groups of the Hamiltonian but which may be used to generate the spectrum, has been adapted from the field of particle physics to the present many-body context. However, unlike the case of particle physics, here we are in possession of a Hamiltonian formulation of the theory, and so it would appear that the method is capable of wide usage.

In the preceding work we have applied the method to the simple case of the reduced Hamiltonian of a superfluid system, thereby obtaining the spectrum and wavefunctions. In this case, the spectrum, at least, is obtainable by other means, but the advantage of the group theory approach, apart from its elegance, is the possibility it provides of a deeper understanding of the mathematical structure of the theory and of a possible comparison with other theories. Thus, while for the superfluid Bose system the relevant group is $SU(1, 1)$ [or $\prod_k \otimes SU(1, 1)_k$] it would appear that the analogous group in the superconductivity case is $SU(2)$ [or $\prod_k \otimes SU(2)_k$].[5] Here the infinite spectrum arises in spite of the fact that we are dealing with a compact group because of the infinite direct product nature of the dynamical group.

In addition, light is thrown on the Bogoliubov transformation, which appears here merely as a rotation in the space of the algebra, and the role played by the sign of the potential in distinguishing between the continuous and discrete spectra.

In principle, the group approach may be used for the solution of any many-body problem involving a Hamiltonian expressed in terms of creation and annihilation operators. In practice, the method is only convenient in those cases where the resulting algebra is of a sufficiently noncomplicated structure.

ACKNOWLEDGMENTS

The author wishes to thank Professor L. O'Raifeartaigh for useful comments, and W. Montgomery for unstinted assistance and critical appraisal.

APPENDIX: THE GROUP $SU(1, 1)$

We review briefly the properties of the noncompact Lie Group $SU(1, 1)$, following Ref. 6. This group is defined as the set of unimodular 2×2 matrices g obeying

$$g\Lambda g^\dagger = \Lambda, \quad \Lambda = \begin{bmatrix} 1 & \\ & -1 \end{bmatrix}. \tag{A1}$$

They may be parametrized by

$$g(\alpha, \beta) = \begin{bmatrix} \alpha & \beta \\ \bar{\beta} & \bar{\alpha} \end{bmatrix}, \quad |\alpha|^2 - |\beta|^2 = 1. \tag{A2}$$

Another parametrization is in terms of three real variables θ_k,

$$g = \exp i\sum_{k=1}^{3}\theta_k J_k, \tag{A3}$$

where the conditions (A1) on g imply that the J_k are traceless and obey

$$J_k\Lambda - \Lambda J_k^\dagger = 0. \tag{A4}$$

A choice of 2×2 matrices J_k to which all other sets are equivalent is

$$J_1 = \frac{1}{2}\begin{bmatrix} & 1 \\ -1 & \end{bmatrix}, \quad J_2 = -\tfrac{1}{2}i\begin{bmatrix} & 1 \\ 1 & \end{bmatrix},$$
$$J_3 = \frac{1}{2}\begin{bmatrix} 1 & \\ & -1 \end{bmatrix}. \tag{A5}$$

These matrices have the commutation relations

$$[J_1, J_2] = -iJ_3, \quad [J_2, J_3] = iJ_1, \quad [J_3, J_1] = iJ_2, \tag{A6}$$

which define the Lie algebra of the Lie group $SU(1, 1)$. We note that the 2×2 matrices (A5) are not all Hermitian; as this algebra is noncompact, there exists no nontrivial finite-dimensional Hermitian representation. It is convenient to introduce the combinations $J^\pm = (\pm J_1 + iJ_2)/\sqrt{2}$ which obey

$$[J_3, J^\pm] = \pm J^\pm, \quad [J^+, J^-] = J_3. \tag{A7}$$

Considering the group as a transformation group on the spinor

$$\xi = \begin{pmatrix}\xi_1 \\ \xi_2\end{pmatrix},$$

thus

$$g: \xi \to \xi' = g(\alpha, \beta)\xi, \tag{A8}$$

we can find representations on the function space of the ξ in the usual way by right translations:

$$g: f \to T_g f, \quad (T_g f)(\xi) = f(\xi g). \tag{A9}$$

Differential operators corresponding to the infinitesimal generators J_k may be obtained by the standard procedure (6):

$$\{J_3, J^+, J^-\} \sim \left\{\frac{1}{2}\left(\xi_1\frac{\partial}{\partial\xi_1} - \xi_2\frac{\partial}{\partial\xi_2}\right), \frac{1}{\sqrt{2}}\xi_1\frac{\partial}{\partial\xi_2}, \frac{1}{\sqrt{2}}\xi_2\frac{\partial}{\partial\xi_1}\right\}. \tag{A10}$$

All the irreducible representations may now be obtained by considering the action of the differential operators (A10) on the monomial $|a, b\rangle = \xi_1^a\xi_2^b$, where a and b may take on arbitrary complex values. Thus

$$J_3|a, b\rangle = \tfrac{1}{2}(a - b)|a, b\rangle, \tag{A11a}$$
$$J^+|a, b\rangle = 2^{-\frac{1}{2}}b\,|a + 1, b - 1\rangle, \tag{A11b}$$
$$J^-|a, b\rangle = 2^{-\frac{1}{2}}a\,|a - 1, b + 1\rangle. \tag{A11c}$$

These equations show that for any irreducible representation the numbers $j = \frac{1}{2}(a + b)$ and $\sigma =$ fractional part of $\frac{1}{2}(a - b)$ are constant. The spectrum of J_3 may be labeled by m, where $\sigma = m + \frac{1}{2}(a - b)$, $m =$ integer. The Casimir operator $C = -J_1^2 - J_2^2 + J_3^2$ obeys

$$C|a, b\rangle = j(j + 1)|a, b\rangle. \tag{A12}$$

The representations $\mathfrak{D}(j, \sigma)$ are irreducible unless one of the coefficients of Eqs. (A11b) and (A11c) vanishes; i.e., either a or b or both are integers. Therefore, the irreducible representations are classified by the following:

$\mathfrak{D}(j, \sigma)$ a,b not integers, choose $-\frac{1}{2} \leq Rl\sigma < \frac{1}{2}$. Spectrum, $J_3 - \sigma = 0, \pm 1, \pm 2, \cdots$. (A13a)

$\mathfrak{D}^+(j, \sigma)$ a integral, b not; $a \geq 0$, $\sigma = -j$. Spectrum, $J_3 - \sigma = 0, 1, 2, \cdots$. (A13b)

$\mathfrak{D}^-(j, \sigma)$ b integral, a not; $b \geq 0$, $\sigma = j$. Spectrum, $J_3 - \sigma = 0, -1, -2, \cdots$. (A13c)

$\mathfrak{D}(j)$ a and b integral, $(a + b) \geq 0$. Finite spectrum $J_3 = \{-j, -j + 1, \cdots, j - 1, j\}$. (A13d)

If, in (A13d), $(a + b) < 0$, then $a \geq 0$ gives $\mathfrak{D}^+$ and $b \geq 0$ gives $\mathfrak{D}^-$ as (A13b) and (A13c).

Unitary irreducible representations are obtained by suitably normalizing the base vectors, $|j, m\rangle = N_m|a, b\rangle$, introducing the inner product

$$\langle j, m\,|\,j, m'\rangle = \delta_{mm'},$$

and imposing the hermiticity condition on the generators,

$$J_k = J_k^\dagger \quad \text{or} \quad (J^+)^\dagger = -J^-.$$

The above representations then become unitary for values of the pair (j, σ) as follows:

$\mathfrak{D}_p(j, \sigma)$, principal series, σ real, $j = -\frac{1}{2} + i\lambda$;

$\mathfrak{D}_s(j, \sigma)$, supplementary series, σ, j real;

$$|j + \tfrac{1}{2}| < \tfrac{1}{2} - |\sigma|. \tag{A14a}$$

$\mathfrak{D}^+(j)$, positive discrete series, real $j = -\sigma$, $j < 0$. (A14b)

$\mathfrak{D}^-(j)$, negative discrete series, real $j = \sigma$, $j < 0$. (A14c)

$\mathfrak{D}(j)$, trivial representation, $j = 0$. (A14d)

The normalization constant for $\mathfrak{D}_p(j, \sigma)$ is given by $N_m = 1$; for all the others

$$N_m = [(m + \sigma - 1 - j)!/(m + \sigma + j)!]^{\frac{1}{2}}.$$

Finally, the finite matrix elements corresponding to the group element $g = g(\alpha, \beta)$ are obtained directly from (A9) as follows:

$$S(g)\,|j, n\rangle = N_n(\alpha\xi_1 + \bar{\beta}\xi_2)^{j+\sigma+n}(\beta\xi_1 + \bar{\alpha}\xi_2)^{j-\sigma-n} = \sum_m S^j_{mn}(g)\,|j, m\rangle.$$

A straightforward calculation gives

$$S^j_{mn}(g) = (N_n/N_m)(j + \sigma + n)!\,(j - \sigma - n)!\,\alpha^{j+\sigma+n}\bar{\alpha}^{j-\sigma-m}\beta^{m-n} \times \sum_k \frac{|\beta/\alpha|^{2k}}{(j + \sigma + n - k)!\,k!\,(j - \sigma - m - k)!\,(k + m - n)!}. \quad \text{(A15)}$$

[1] C. Fronsdal, Phys. Rev. 156, 1666 (1967).
[2] Y. Dothan, M. Gell-Mann, and Y. Ne'eman, Phys. Letters 17, 283 (1965).
[3] W. H. Bassichis and L. L. Foldy, Phys. Rev. 133, A935 (1964).
[4] N. N. Bogoliubov, J. Phys. (USSR) 11, 23 (1947).
[5] The treatment of P. W. Anderson [Phys. Rev. 112, 900 (1958)], by analogy with a system of spins, suggests this.
[6] A. O. Barut and C. Fronsdal, Proc. Roy. Soc. (London) 287A, 532 (1965).
[7] See, for example, M. Hamermesh, *Group Theory* (Addison-Wesley, Reading, Mass., 1962), pp. 296f.

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 12, NUMBER 3 MARCH 1971

Topology for Minkowski Space

S. NANDA
Mathematics Department, Regional Engineering College, Rourkela-8, Orissa, India

(Received 5 March 1970)

The purpose of this paper is to prove one of Zeeman's conjectures that the group of homeomorphisms of the finest topology on Minkowski space that induces the 3-dimensional Euclidean topology on spacelike hyperplanes is the one generated by the inhomogeneous Lorentz group and dilatations.

1. INTRODUCTION

Apart from the pure mathematical interest, there are several reasons for finding out topologies for Minkowski space, the 4-dimensional space–time continuum of special relativity, whose homeomorphism groups form precisely the inhomogeneous Lorentz group. Since Minkowski space is a 4-dimensional real vector space, it is natural to consider the Euclidean topology which is generated by the usual Euclidean distance function. Such a topology is locally homogeneous and hence unsuitable, because, at each point of Minkowski space, the null cone separates the spacelike vectors from the timelike ones. Moreover, the group of homeomorphisms in the case of Euclidean topology is too large to be of any physical significance. For example, one can define a linear map which maps a spacelike vector to a timelike vector; this is clearly a homeomorphism with respect to the Euclidean topology. This topology is completely unrelated to the quadratic form, which is physically so important.

As is well known, the indefinite quadratic form on Minkowski space, in a very natural manner, gives rise to a causal relation, i.e., a relation of precedence, which in the mathematical sense is simply a partial order on the collection of space–time events. Thus, as a partially ordered set, Minkowski space can be given the "interval topology,"[1] but the interval topology and the Euclidean topology coincide in the case of Minkowski space.[2] Zeeman[3] suggested that Minkowski space be given a topological structure appropriate to the mathematical structure of the space, that is, a topology which "fits" the indefinite fundamental form and the null cones associated with it.

The topology suggested by Zeeman (which he calls the "fine topology") is defined as the finest topology on Minkowski space such that the induced topology on every spacelike hyperplane and every timelike line

The normalization constant for $\mathfrak{D}_p(j, \sigma)$ is given by $N_m = 1$; for all the others

$$N_m = [(m + \sigma - 1 - j)!/(m + \sigma + j)!]^{\frac{1}{2}}.$$

Finally, the finite matrix elements corresponding to the group element $g = g(\alpha, \beta)$ are obtained directly from (A9) as follows:

$$S(g)\,|j, n\rangle = N_n(\alpha\xi_1 + \bar{\beta}\xi_2)^{j+\sigma+n}(\beta\xi_1 + \bar{\alpha}\xi_2)^{j-\sigma-n} = \sum_m S^j_{mn}(g)\,|j, m\rangle.$$

A straightforward calculation gives

$$S^j_{mn}(g) = (N_n/N_m)(j + \sigma + n)!\,(j - \sigma - n)!\,\alpha^{j+\sigma+n}\bar{\alpha}^{j-\sigma-m}\beta^{m-n} \times \sum_k \frac{|\beta/\alpha|^{2k}}{(j + \sigma + n - k)!\,k!\,(j - \sigma - m - k)!\,(k + m - n)!}. \quad \text{(A15)}$$

[1] C. Fronsdal, Phys. Rev. 156, 1666 (1967).
[2] Y. Dothan, M. Gell-Mann, and Y. Ne'eman, Phys. Letters 17, 283 (1965).
[3] W. H. Bassichis and L. L. Foldy, Phys. Rev. 133, A935 (1964).
[4] N. N. Bogoliubov, J. Phys. (USSR) 11, 23 (1947).
[5] The treatment of P. W. Anderson [Phys. Rev. 112, 900 (1958)], by analogy with a system of spins, suggests this.
[6] A. O. Barut and C. Fronsdal, Proc. Roy. Soc. (London) 287A, 532 (1965).
[7] See, for example, M. Hamermesh, *Group Theory* (Addison-Wesley, Reading, Mass., 1962), pp. 296f.

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 12, NUMBER 3 MARCH 1971

Topology for Minkowski Space

S. NANDA
Mathematics Department, Regional Engineering College, Rourkela-8, Orissa, India

(Received 5 March 1970)

The purpose of this paper is to prove one of Zeeman's conjectures that the group of homeomorphisms of the finest topology on Minkowski space that induces the 3-dimensional Euclidean topology on spacelike hyperplanes is the one generated by the inhomogeneous Lorentz group and dilatations.

1. INTRODUCTION

Apart from the pure mathematical interest, there are several reasons for finding out topologies for Minkowski space, the 4-dimensional space–time continuum of special relativity, whose homeomorphism groups form precisely the inhomogeneous Lorentz group. Since Minkowski space is a 4-dimensional real vector space, it is natural to consider the Euclidean topology which is generated by the usual Euclidean distance function. Such a topology is locally homogeneous and hence unsuitable, because, at each point of Minkowski space, the null cone separates the spacelike vectors from the timelike ones. Moreover, the group of homeomorphisms in the case of Euclidean topology is too large to be of any physical significance. For example, one can define a linear map which maps a spacelike vector to a timelike vector; this is clearly a homeomorphism with respect to the Euclidean topology. This topology is completely unrelated to the quadratic form, which is physically so important.

As is well known, the indefinite quadratic form on Minkowski space, in a very natural manner, gives rise to a causal relation, i.e., a relation of precedence, which in the mathematical sense is simply a partial order on the collection of space–time events. Thus, as a partially ordered set, Minkowski space can be given the "interval topology,"[1] but the interval topology and the Euclidean topology coincide in the case of Minkowski space.[2] Zeeman[3] suggested that Minkowski space be given a topological structure appropriate to the mathematical structure of the space, that is, a topology which "fits" the indefinite fundamental form and the null cones associated with it.

The topology suggested by Zeeman (which he calls the "fine topology") is defined as the finest topology on Minkowski space such that the induced topology on every spacelike hyperplane and every timelike line

is Euclidean. The fine topology is physically significant in the sense that the group of its homeomorphisms is precisely the inhomogeneous Lorentz group together with dilatations. It turns out, on investigation, that there exists a wide class of topologies on Minkowski space each of which has the same homeomorphism group as that of the fine topology. The general procedure for defining such topologies is to prescribe the induced topologies on spacelike hyperplanes and timelike and lightlike lines. In the case of the fine topology, one has information about two types of compact sets, namely the 3-dimensional ones on spacelike hyperplanes and the 1-dimensional ones on timelike lines. This is the main reason why it is possible to derive the homeomorphism group with considerable simplicity. If, however, we have only one piece of information, i.e., either about compact sets on spacelike hyperplanes or about compact sets on timelike lines, then it becomes much more difficult to derive the homeomorphism group and, in certain cases, an almost impossible task. The object of this paper is to consider one such case and to prove one of Zeeman's conjectures[3] that the finest topology on Minkowski space that induces Euclidean topology on spacelike hyperplanes has the same group of homeomorphisms as that of the fine topology.

Not many results of topology and analysis are needed to prove the conjecture. Only elementary notions of topology such as "compactness," "connectedness," and "limit point," etc., are used throughout. Baire's category theorem and elementary properties of Lebesgue measure on the real line are about the only results from analysis which have been employed in the course of the proof.

2. NOTATION AND TERMINOLOGY

Let M denote Minkowski space with characteristic quadratic form Q:

$$M = \{(x_0, x_1, x_2, x_3): x_i \text{ are reals}\},$$
$$Q(x) = x_0^2 - x_1^2 - x_2^2 - x_3^2.$$

Let

$$K = \{x \in M: Q(x) > 0 \text{ and } x_0 > 0\},$$
$$L = \{x \in M: Q(x) \geq 0 \text{ and } x_0 > 0\}.$$

It is easy to verify that K satisfies the following three conditions:

(i) $K + K = \{x + y: x \in K, y \in K\} \subset K$;
(ii) $aK = \{ax: x \in K\} \subset K$ for every positive number a;
(iii) $K \cap (-K) = \phi$, where $-K = \{-x: x \in K\}$.

L also satisfies (i), (ii), and (iii). K and L are so-called positive cones, and each generates a partial order on M as follows:

$$x < y \Leftrightarrow y - x \in K,$$
$$x \ll y \Leftrightarrow y - x \in L.$$

There is also another relation $<.$ on M defined by $x <. y \Leftrightarrow Q(y - x) = 0$ and $y_0 - x_0 > 0$, i.e., $y - x$ is a future-pointing lightlike (null) vector; this is not a partial order due to the lack of transitivity. With this notation, it is easy to establish that

$$x \ll y \Leftrightarrow \begin{cases} \text{either } x < y \\ \text{or } x <. y. \end{cases}$$

A real vector space with a partial order $<$ is said to be a partially ordered vector space (POVS), if the order structure is compatible with the vector space structure, i.e., (i) $x > 0$, $y > 0 \Rightarrow x + y > 0$ and (ii) $x > 0$, $a > 0 \Rightarrow ax > 0$, where a is a real number. From the properties of the positive cones K and L, it follows easily that M is a POVS with respect to $<$ or $\ll$.

A mapping f of $(M, <)$ into itself is said to be order-preserving if $x < y \Rightarrow f(x) < f(y)$. If f is one to one, then it is said to be inverse-order-preserving if f^{-1} is order-preserving. A one-to-one mapping of M onto itself which is both order-preserving and inverse-order-preserving is called an automorphism of M. Since we have two partial orders $<$ and $\ll$, we shall write $<$-automorphism or $\ll$-automorphism depending on which order we use. Similarly, we shall use the words $<$-preserving and $\ll$-preserving, etc.

Two elements x and y of M are said to be comparable with respect to the partial order $<$ (respectively, with respect to $\ll$) if either $x < y$ or $y < x$ (respectively $x \ll y$ or $y \ll x$). A subset P of M is said to be linearly ordered with respect to $<$ (respectively $\ll$) if any two elements of P are comparable with respect to $<$ (respectively $\ll$). Let G be the group of one-to-one mappings of M onto itself consisting of (i) the Lorentz group, i.e., all linear maps which leave the quadratic form Q invariant, (ii) translations, and (iii) dilatations. Let G_0 be the subgroup of G consisting of the $<$-automorphisms of M. Since every element of G either preserves or reverses the partial order $<$ in M, it follows that G_0 is of index 2 in G. Zeeman[4] has proved that the group of $<$-automorphisms of M is G_0. (This result will be referred to as Zeeman's theorem hereafter.)

We have the following cones at x:

light cone at x: $C^{\mathrm{L}}(x) = \{y: Q(y - x) = 0\}$,
time cone at x: $C^{\mathrm{T}}(x) = \{y: Q(y - x) > 0\} \cup \{x\}$,
space cone at x: $C^{\mathrm{S}}(x) = \{y: Q(y - x) < 0\} \cup \{x\}$,
$$C^{\mathrm{LT}}(x) = C^{\mathrm{L}}(x) \cup C^{\mathrm{T}}(x).$$

It may be easily seen that if $K^*(x) = (K + x) \cup \{x\}$ and $L^*(x) = (L + x) \cup \{x\}$, then $C^{\mathrm{T}}(x) = K^*(x) \cup [-K^*(x)]$ and similarly $C^{\mathrm{LT}}(x) = L^*(x) \cup [-L^*(x)]$. Since every element of G leaves the sign of Q fixed, it is clear that all these cones are invariant under G. With the above notation, it is also easy to see that if f is a $<$-automorphism, then $f(K^*(x)) = K^*(fx)$ and, similarly, if f is a $\ll$-automorphism, then $f(L^*(x)) = L^*(fx)$.

3. $\ll$-AUTOMORPHISMS OF M

In this section, we prove that the group of $\ll$-automorphisms of M is G_0. We start by proving a lemma.

Lemma 1: Let f be a $\ll$-automorphism and $x <. y$; then $fx <. fy$.

Proof: Let $[x, y]$ denote the closed line segment joining x and y. Since $x <. y$, one has $[x, y] = L^*(x) \cap [-L^*(y)]$, and, since f is a $\ll$-automorphism, it follows that

$$f[x, y] = f(L^*(x) \cap [-L^*(y)]) = L^*(fx) \cap [-L^*(fy)].$$

Now suppose to the contrary that $fx <. fy$ is not true; then either (i) $fx < fy$ or (ii) $fx \ll fy$ is false. In the second case, $L^*(fx) \cap [-L^*(fy)] = \phi = f[x, y]$, which is a contradiction, since f is a one-to-one map. In the first case, it is possible to choose a Euclidean open set O in $L^*(fx) \cap [-L^*(fy)]$ and two points p and q in O such that neither $p \ll q$ nor $q \ll p$. But $[x, y]$ is a linearly ordered set with respect to $\ll$, and so is $f[x, y]$, since f is a $\ll$-automorphism. Thus we have a contradiction, and the lemma is proved.

Theorem 1: The group of $\ll$-automorphisms of M is G_0.

Proof: It follows from Lemma 1 that the forward and backward light cones are preserved by any $\ll$-automorphism; therefore, $f(K(x)) = K(fx)$. Thus, f is a $<$-automorphism of M and, by Zeeman's theorem, is an element of G_0. On the other hand, every element of G_0 is clearly a $\ll$-automorphism. This completes the proof.

4. DEFINITION AND PROPERTIES OF SPACE TOPOLOGY

Definition 1: The space topology on M is defined as the finest topology with respect to which the induced topology on every spacelike hyperplane is Euclidean.

Denote by M^{S} and M^{E} the set M equipped with the space topology and the Euclidean topology, respectively.

By definition, the space topology is finer than the Euclidean topology and hence Hausdorff.

Let

$$N_\epsilon^{\mathrm{E}}(x) = \{y : d(x, y) < \epsilon, \text{ where } d \text{ is the Euclidean distance function on } M \text{ and } \epsilon > 0\}$$

and let $N_\epsilon^s(x) = N_\epsilon^{\mathrm{E}}(x) \cap C^{\mathrm{S}}(x)$. It is easy to establish that the topology generated by the sets $\{N_\epsilon^s(x)\}$ induces the 3-dimensional Euclidean topology on every spacelike hyperplane. If we call this topology the s-topology, then the space topology, by definition, is at least as fine as the s-topology. Indeed, the following example will establish that the space topology is strictly finer than the s-topology.

Let $\{H_n\}$ be a sequence of distinct spacelike hyperplanes passing through a point z. Choose a point $z_n \in H_n$ such that $d(z_n, z) \to 0$ as $n \to \infty$ and, moreover, such that not more than a finite number of z_n should be contained in any H_n. Let $Z = \{z_n\}$. We shall now show that Z is closed in M^{S}; it is enough to prove that $H \cap Z$ is a finite set, where H is any spacelike hyperplane. Suppose to the contrary that $H \cap Z$ is infinite; then, H being complete in the induced Euclidean topology, the sequence $H \cap Z$ must converge to a point of H, and, since H is Hausdorff, the point must be z. Thus H passes through the point z, and, by our choice, $H \cap Z$ is at most finite. This is a contradiction and the assertion is proved. Z^{c} is then open in M^{S} (where Z^{c} denotes the complement of the set Z). On the other hand, Z^{c} is not open in the s-topology since any neighborhood $N_\epsilon^s(z)$ of z in this topology will meet Z.

The space topology induces the discrete topology on every timelike or lightlike line, for, if λ is any such line, then, for any $x \in \lambda$, $\lambda \cap N_\epsilon^s(x) = \{x\}$. Following either a measure-theoretic argument or simple topological arguments, we can also establish that the space topology is not normal. It is not locally compact, nor does it have a countable base of neighborhoods.

5. ZENO SEQUENCE

Definition 2: A Zeno sequence $Z = \{z_n\}$ in M^{S} is a sequence of distinct points of M not containing z such that $z_n \to z$ in M^{E} and $z_n \not\to z$ in M^{S}.

$z_n \to z$ in M^{E} implies that every ϵ-neighborhood of z in M^{E} meets Z and therefore $z \in \bar{Z}^{\mathrm{E}}$ (where the bar followed by E indicates closure in the Euclidean topology). Since z is not in Z, it follows that Z is not closed in M^{E}. On the other hand, $z_n \not\to z$ in M^{S} means

that there exists a neighborhood O of z in the space topology such that $O \cap Z = \phi$; hence z is an interior point of Z^c in M^S. We claim that Z^c is open in M^S. To prove this, it is enough to show that every $x \neq z$ of Z^c is an interior point of Z^c in M^S. Let $x \neq z$ be any point of Z^c; then there exists a Euclidean neighborhood of x not meeting Z (which also serves as neighborhood in the space topology since the space topology is finer than the Euclidean topology); for, if not, then every Euclidean neighborhood of x will contain infinitely many points of Z, and, consequently, x will be a limit point of Z in M^E, thus giving a contradiction (since M^E is Hausdorff and $x \neq z$). Thus, Z^c is open in M^S, and consequently Z is closed.

Conversely, if we have a sequence $Z = \{z_n\}$ of distinct points such that Z is closed in M^S and not closed in M^E, then Z must be a Zeno sequence by definition. Thus, we have proved the following.

Proposition 1: A sequence $Z = \{z_n\}$ of distinct points of M^S not containing z is a Zeno sequence iff Z is closed in M^S and not closed in M^E.

In view of the peculiarities of the space topology, it is easy to construct Zeno sequences.

Example 1: Consider the sequence $Z = \{z_n\}$ as in Sec. 4 above; Z is closed in M^S and not closed in M^E; it is therefore a Zeno sequence.

Example 2: Let $\{t_n\}$ be a sequence of distinct timelike or lightlike (or a combination of both) lines passing through a point z. Choose $z_n \in t_n$ such that $d(z_n, z) \to 0$ as $n \to \infty$; then, clearly, $z_n \to z$ in M^E. On the other hand, a neighborhood of the form $N_\epsilon^s(z)$ does not meet $Z = \{z_n\}$; hence $z_n \not\to z$ in M^S. Z is thus a Zeno sequence.

The most important property of a Zeno sequence that will be used subsequently is the following.

Lemma 2: A compact set of M^S cannot contain a Zeno sequence.

Proof: Let D be a compact set of M^S containing a Zeno sequence Z. Z being a closed subset of a compact set is compact. It is easy to see that the topology induced by the space topology on Z is discrete because one can always choose a Euclidean ϵ-neighborhood about any point $z_n \in Z$ which contains no other point of Z. Since the points of Z are distinct, it follows that Z is an infinite discrete set and therefore cannot be compact; thus we have a contradiction.

6. HOMEOMORPHISMS OF M^S

This section is devoted to the derivation of the group of homeomorphisms of M^S. We now go through a series of lemmas.

Lemma 3: Let h be a homeomorphism of M^S and B a 3-dimensional closed ball with center x, contained in a spacelike hyperplane H passing through x; then there exists a ball B_r of radius r and center x such that (i) $B_r \subset B$ and (ii) $hB_r \subset C^S(hx)$.

Proof: First, observe that hB cannot be contained entirely in $C^{LT}(hx)$; for, if this is the case, then $hB \cap N_\epsilon^s(hx) = \{hx\}$ is an open set in hB; i.e., $\{x\}$ is an open set in B which is false since the induced topology on B is Euclidean, and hence every singleton is closed.

Thus there are two possibilities: (i) either $hB \subset C^S(hx)$ or (ii) hB is contained partly in $C^S(hx)$ and partly in $C^{LT}(hx)$. In case (i), take $B = B_r$. In case (ii), $C^S(hx) \cap hB$ is an open subset of hB containing hx; hence, its inverse image, D say, is an open subset of B, and clearly $hD \subset C^S(hx)$. Since B has Euclidean topology, it is possible to choose a ball B_r of radius r such that $B_r \subset D \subset B$. Thus, in either case, the lemma is true.

Since B_r is compact and connected, so is the image hB_r. We shall now examine more closely the nature of the image hB_r. The following two lemmas will give a complete description of this set.

Lemma 4: Let B_r and h be as in Lemma 3; then hB_r is contained in the union of a finite number of spacelike hyperplanes.

Proof: First, we claim that it is possible to choose $\epsilon > 0$ such that $N_\epsilon^s(hx) \cap hB_r$ is contained in the union of a finite number of spacelike hyperplanes through hx. Suppose to the contrary that this is not possible; then, for every $\epsilon > 0$, $N_\epsilon^s(hx) \cap hB_r$ meets an infinite number of spacelike hyperplanes through hx. One can then construct a Zeno sequence in hB_r as in Sec. 4. Since B_r is compact, this is a contradiction in view of Lemma 2, and our assertion is proved.

Choose ϵ as above; then $h^{-1}[N_\epsilon^s(hx)] \cap B_r = O_x$ is an open set about x in the induced topology of B_r; moreover, from the choice of ϵ above, it follows that $h(O_x)$ is contained in a finite union of spacelike hyperplanes through hx. Since the construction given above is valid for each x in B_r (i.e., for each x in B_r, it is possible to choose an open set O_x such that hO_x is contained in a finite union of spacelike hyperplanes through hx) and B_r is compact, the proof is complete.

Lemma 5: Let H be a spacelike hyperplane through x and h a homeomorphism of M^S; then $M^E - hH$ is disconnected and has two components.

Proof: Note that the induced topology on H is 3-dimensional Euclidean, so that hH is topologically R^3. Moreover, $M^S - H$ has two components; hence $M^S - hH$ also has two components. Since the space topology is finer than the Euclidean topology, it follows that $M^E - hH$ has at most two components. We claim that $M^E - hH$ has exactly two components; for, if not, then $M^E - hH$ is connected, and therefore any two points p and q of $M^E - hH$ can be arc-connected by a polygonal path which is piecewise spacelike. (This is possible since hH is closed; hence $M^E - hH$ is open in M^E, and every open connected subset of M^E is arc-connected.) Since the topology induced by the space topology on such a polygonal path is Euclidean, it follows that $M^S - hH$ is arcwise connected and therefore connected, which is a contradiction. Thus, the proof of the lemma is complete.

Lemma 6: Let w be a spacelike interval passing through a point x, and let h be a homeomorphism of M^S; then there exists a point z in w such that $h[x, z] = [hx, hz]$, where $hz \in C^S(hx)$.

Proof: For convenience, let us first consider the corresponding case in a 3-dimensional Minkowski space. Let H_1 and H_2 be two spacelike planes whose intersection contains the spacelike interval w. If we choose $B_1 \subset H_1$ and $B_2 \subset H_2$ as in Lemma 3, then hB_1 and hB_2 are each contained in a finite number of spacelike planes, say m and n planes, respectively. Suppose that $H_1, H_2, \cdots, H_m$ are the first m planes and $H_1', H_2', \cdots, H_n'$ are the second n planes; then

$$(hB_1 \cap hB_2) \subset \left[\left(\bigcup_{i=1}^{m} H_i\right) \cap \left(\bigcup_{k=1}^{n} H_k'\right)\right].$$

Observe that the intersection of two planes is a straight line, so that the right-hand side of the above inclusion relation is a union of at most mn straight lines, each of which is spacelike. Moreover, $B_1 \cap B_2$ is connected; hence $hB_1 \cap hB_2$ is a connected subset of those mn spacelike lines. It is enough to take a point z' in one linear segment starting from hx such that $z' \in hw$ and to take $z = h^{-1}z'$. This clearly serves the purpose of the lemma. However, another point which is crucial in the above proof must be taken care of. One has to consider the case when one of the planes H_i is the same as one of the planes H_k'. If $S = (H_i \cap hB_1) \cap (H_k' \cap hB_2) \neq \phi$, then, clearly, it is possible to choose a set in S which is open in the induced topology of either hB_1 or hB_2. This means that one can choose a set in $B_1 \cap B_2$ which is open in the induced topology of either B_1 or B_2. This is not possible since $B_1 \cap B_2$ is a linear interval. This completes the proof in case of the 3-dimensional Minkowski space.

One can now apply similar arguments to the 4-dimensional case. Consider two spacelike hyperplanes H_1 and H_2 passing through x such that $w \subset H_1 \cap H_2$. Choose $B_1 \subset H_1$ and $B_2 \subset H_2$ as in Lemma 3. The foregoing argument then shows that $hB_1 \cap hB_2$ is a subset of a finite union of spacelike 2-planes, whereas $B_1 \cap B_2 = s_1$ is itself a subset of a spacelike 2-plane. Take another spacelike 2-plane passing through x, and choose a subset s_2 whose image is again a subset of a finite union of spacelike 2-planes. Without loss of any generality, take $w \subset s_1 \cap s_2$; then hw is a finite union of spacelike intervals. Note that w is connected in the induced Euclidean topology, so that hw is a finite union of piecewise linear spacelike intervals. Choose a point z' in the first piece; thus $z = h^{-1}(z')$ serves the purpose of the lemma, and this completes the proof.

Our next object is to prove that the partial order $\ll$ is preserved at least locally. To do this, we need the following lemma:

Lemma 7: Let R be the real line with the usual topology and $f: R \to R$ a strictly positive real-valued function on R; then, for almost every x (i.e., for every x in $R - S$, where S is a set of measure zero), there exists a subset D_x in R satsifying the following conditions: (i) $x \in \text{Int}(\bar{D}_x)$, where Int denotes "interior" and the bar denotes the closure of a set, and (ii) $f(y) > k_x > 0$ for every $y \in D_x$ for some constant k_x.

Remark 1: "Measure" means the usual Lebesgue measure on the real line.

Proof: Let $A_n = \{x \in R: f(x) > 1/n\}$. Since $f(x) > 0$ for every $x \in R$, we have $\bigcup_{n=1}^{\infty} A_n = R$. It follows now from Baire's category theorem (which states that a complete metric space cannot be expressed as a countable union of nowhere dense sets) that at least one of the sets A_n is not nowhere dense. Suppose that A_k is such a set; then $\text{Int}(\bar{A}_k) \neq \phi$ and $f(y) > 1/k$ for every $y \in A_k$. For every $x \in \text{Int}(\bar{A}_k)$, if we choose $D_x = A_k$ and $k_x = 1/k$, then clearly the two conditions of the lemma are satisfied. Note that $R - \text{Int}(\bar{A}_k)$ is again a complete metric space and hence belongs to the second category, so that we can apply the preceding argument again to $R - \text{Int}(\bar{A}_k)$. Let $P = \{x \in R:$ there exists a set D_x as above satisfying conditions

(i) and (ii) of the lemma}. Since $\bigcup_{x\in I'} \mathrm{Int}\,(\bar{D}_x)$ is an open set, $S = R - P$ is closed. We claim that $m(S) = 0$, where m denotes the Lebesgue measure on the real line. Suppose to the contrary that $m(S) \neq 0$; then S must contain an interval I. (A closed set of nonzero measure must contain an interval.[5]) Choose a closed interval $C \subset I$. Since C is a complete metric space, it is possible to apply the same construction as above to get a point x in C having a corresponding set D_x and satisfying conditions (i) and (ii) of the lemma; thus $x \in P$. This is a contradiction, and the lemma is proved.

Remark 2: Note that the set S which is of measure zero is closed. We call it the exceptional set.

Remark 3: If R is a line with some other topology, then the preceding lemma is valid only when Int and closure refer to the usual Euclidean topology of the line.

Remark 3 is significant, since we want to apply the lemma to a timelike or a lightlike line in M^S whose induced topology is discrete.

In the following lemma, we shall assume, without loss of generality, that D_x is everywhere dense in $\mathrm{Int}^E\,(\bar{D}_x^E)$, where E refers to the Euclidean topology of the line.

Lemma 8: Let T be a timelike or a lightlike line and h a homeomorphism of M^S; then almost everywhere on T (i.e., except on a set S of measure zero, which is Euclidean-closed in this case) the following property is satisfied: If $x \in T - S$, then there is a point $z \gg x$ in T such that $h[x, z] = [hx, hz]$ with $hx \ll hz$ or $hz \ll hx$.

Proof: Let w and w' be two distinct (fixed) spacelike directions such that the 2-plane spanned by w and w' is spacelike. For each x in T, let w_x and w'_x be the half-lines starting from x and parallel to w and w', respectively. By Lemma 6, there are points z and z' on w and w', respectively, such that $[x, z]$ and $[x, z']$ are mapped by h to spacelike intervals. If $f(x) = \min\,(d(x, z), d(x, z'))$, then $f(x) > 0$ for every x in T, and Lemma 7 applies; i.e., for each $x \in T - S$, where S is a Euclidean closed subset of T of measure zero, we have a set D_x such that $x \in \mathrm{Int}^E\,(\bar{D}_x^E)$ and $f(y) > k_x > 0$ for every $y \in D_x$.

Let $x \in T - S$ and D_x be as above. The directions T (future-pointing) and w determine a 2-plane s, and similarly the direction T (future-pointing) and w' determine another 2-plane s'. We now confine our attention to the 2-plane s. (See Fig. 1.)

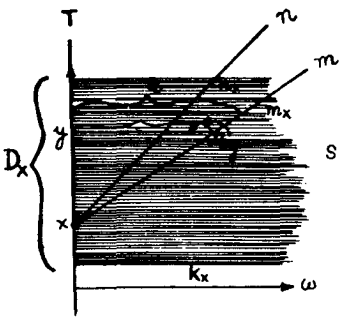

FIG. 1. Construction of $\bar{\Delta}$.

Take two spacelike half-lines m and n contained in the 2-plane s and starting from the point x. By Lemma 6, there are points m_x and n_x in m and n, respectively, such that $[x, m_x]$ and $[x, n_x]$ are mapped by h to spacelike intervals. Note that $[x, m_x]$ and $[x, n_x]$ meet an infinite number of spacelike intervals starting from the points of D_x which are parallel to w and whose images by h are intervals. This is obvious in view of the fact that, for each y in D_x, $f(y) > k_x > 0$ and D_x is dense in $\mathrm{Int}^E\,(\bar{D}_x^E)$. Clearly, these meeting points are dense in $[x, m_x]$ and $[x, n_x]$, whose induced topology is Euclidean. Let q be a spacelike interval parallel to w which meets $[x, m_x]$ and $[x, n_x]$ and meets T at a point of D_x. Let Δ be the union of closed linear spacelike intervals starting from points of D_x, parallel to w, whose images are spacelike intervals and which are bounded by the straight lines T, m (or n as the case may be), and the interval q. Since the topology induced on spacelike intervals is Euclidean and since these intervals meet $[x, m_x]$ and $[x, n_x]$ in dense sets, it follows that $\bar{\Delta}$ (closure taken in the space topology) is a triangular area bounded by the above three directions.

Now consider the image $h(\bar{\Delta})$. Recall that $h[x, m_x]$ and $h[x, n_x]$ are spacelike linear intervals by choice. If q_y is the closed spacelike linear interval parallel to q, starting from $y \in D_x$ and ending with a point on m, then each q_y is mapped by h to a spacelike linear interval (by definition of the function f on T). Thus $h(\bar{\Delta})$ is a set contained in a 2-plane which is completely determined by the intervals $h[x, m_x]$, $h[x, n_x]$, and $h(q)$.

In a similar manner, if we consider the 2-plane s' determined by T and w', then we get a triangular area $\bar{\Delta}'$ as above and its corresponding image $h(\bar{\Delta}')$. Observe that $\bar{\Delta} \cap \bar{\Delta}' = I$ is an interval. On the other hand, since $h(\bar{\Delta})$ and $h(\bar{\Delta}')$ are each contained in a 2-plane, it follows that $hI = h(\bar{\Delta} \cap \bar{\Delta}') = h(\bar{\Delta}) \cap h(\bar{\Delta}')$ is a subset of a straight line. We assert that hI is connected in the Euclidean topology, i.e., hI is an interval.

Suppose to the contrary that hI is disconnected in the Euclidean topology; then there exists a spacelike

hyperplane H such that (i) $H \cap hI = \phi$ and (ii) both components of $M^S - H$ have nonvoid intersection with hI. On the other hand, in view of Lemma 5, the inverse image $h^{-1}H$ is such that only one component of $M^E - h^{-1}H$ contains points of I. This is a contradiction, and our assertion is proved, i.e., hI is connected in the Euclidean sense; in other words, hI is an interval. To see whether hI is spacelike, timelike, or lightlike is not difficult. Since I is discrete in its induced topology, so is hI; therefore, hI is either timelike or lightlike. Now choose $z \in I$ with $x \ll z$, and the proof of the lemma is complete.

Remark 4: In a similar way, one can choose a point $z' \ll x$ such that $h[x, z'] = [hx, hz']$ with $hx \ll hz'$ or $hz' \ll hx$.

It is necessary at this point to observe that we are confining our attention to only one timelike or lightlike line and particularly to the part $T - S$, where S is the exceptional set of measure zero. Since S is closed in the Euclidean topology of the line, $T - S$ is open and therefore a countable union of Euclidean open intervals. Let O be one of those intervals. We shall show in the following lemma that at any point x of O, the partial order $\ll$ is locally preserved or reversed.

Lemma 9: Let x be a point of O, and let z and z' be determined as in Lemma 8 with $z' \ll x \ll z$; i.e., z and z' are oppositely oriented with respect to hx; then hz and hz' are oppositely oriented with respect to hx.

Proof: Suppose to the contrary that hz and hz' are oriented in the same way with respect to hx, and suppose, without loss of generality, that $hx \ll hz$ and $hx \ll hz'$. From Lemma 8, we have $h[x, z] = [hx, hz]$ and $h[x, z'] = [hx, hz']$. Consider a spacelike hyperplane H passing through hx. Clearly, $M^S - H$ is disconnected, having two components, and $(hx, hz]$ and $(hx, hz']$ are contained in the same component, where $(hx, hz]$ denotes the open closed interval from hx to hz. On the other hand, in the inverse image, $(x, z]$ and $(x, z']$ are contained in different components (via Lemma 5). Thus, we have a contradiction, and the lemma is proved.

Corollary 1: Since the preceding argument applies to any point of the interval O, it follows that, at any point O, h is either $\ll$-order preserving or $\ll$-order-reversing. In other words, if h preserves (or reverses) the order $\ll$ at any point of O, then it does so at every point of O. This is a direct consequence of the fact that any two points of O can be contained in a subset C which is compact in the Euclidean topology.

The next important step will be to examine what happens to the points of the exceptional set S, i.e., to see how these points are mapped by a homeomorphism and whether at such points the order relation $\ll$ is locally preserved or reversed.

Lemma 10: Let p be a point of the exceptional set S such that it is the end point of an interval O (as described at the end of the proof of Lemma 8), and let q be a point of O; then $h[p, q]$ is connected in the Euclidean topology.

Proof: It is already known from Lemma 8 that $h(p, q]$ is connected in the Euclidean topology. Suppose that $h[p, q]$ is disconnected in the Euclidean topology; then it is possible to disconnect the space M either by removing a spacelike hyperplane H or by removing H', which is the homeomorph of a spacelike hyperplane, such that $H \cap h[p, q] = \phi$ (or $H' \cap h[p, q] = \phi$). Moreover, hp and $h(p, q]$ are contained in different components of $M^S - H$ (or $M^S - H'$). On the other hand, in the inverse image, it is impossible to disconnect M^S by removing a homeomorph of R^3 which does not meet $[p, q]$ such that p and $(p, q]$ will be in different components. The contradiction proves the lemma.

The following lemma will show that Lemma 9 is also valid for points of S.

Lemma 11: Let $x \in S$ be the end point of two Euclidean open intervals O_1 and O_2 (where O_1 and O_2 are subsets of $T - S$). Let $z_1 \in O_1$ and $z_2 \in O_2$ be such that $z_2 \ll x \ll z_1$, i.e., z_1 and z_2 are oppositely oriented with respect to x; then hz_1 and hz_2 are oppositely oriented with respect to hx.

Proof: Suppose to the contrary that hz_1 and hz_2 are oriented in the same way, i.e., assume without loss of generality that $hx \ll hz_1$ and $hx \ll hz_2$. We already know from Lemma 8 that $h(x, z_1]$ and $h(x, z_2]$ are piecewise linear and, from Corollary 1, that h either preserves or reverses the order of $(x, z_1]$ and $(x, z_2]$. Now take a spacelike hyperplane H passing through hx. In view of Lemma 10, it is clear that $M^S - H$ has two components, one containing $h(x, z_1]$ and $h(x, z_2]$. On the other hand, in the inverse image, $(x, z_1]$ and $(x, z_2]$ are contained in different components of $M^S - h^{-1}H$. This is a contradiction, and the lemma is proved.

Lemmas 9 and 11 together prove that if T is a timelike or a lightlike line and h a homeomorphism of M^S, then $h \mid T$, i.e., h restricted to T, is either $\ll$-order-preserving or $\ll$-order-reversing. It now remains to be shown that this remains true globally.

Lemma 12: Let T_1 and T_2 be two timelike or (and) lightlike lines and h a homeomorphism of M^S; then T_1 and T_2 are mapped with the same orientation.

Proof: Suppose to the contrary that h preserves the orientation of T_1 and reverses that of T_2. First, assume that T_1 and T_2 intersect at a point x. Take a spacelike hyperplane H at x; then $M^S - H$ has two components, one containing T_1' and T_2' and the other containing T_1'' and T_2'', where $\{x\} \cup T_1' \cup T_1'' = T_1$ and $\{x\} \cup T_2' \cup T_2'' = T_2$. On the other hand, $M^S - hH$ has two components, one containing hT_1'' and hT_2' and the other containing hT_2'' and hT_1'. This is a contradiction. Secondly, if T_1 and T_2 are disjoint, then take a third timelike or lightlike line T_3 which meets both T_1 and T_2. Let $T_1 \cap T_3 = \{x\}$ and $T_2 \cap T_3 = \{y\}$. By applying the argument of the first case at x, one concludes that h preserves the orientation of T_3, and, similarly applying it at y, one concludes that orientation of T_2 is preserved, which is a contradiction. Thus we have proved that the order relation $\ll$ is either preserved or reversed in M.

Theorem 2: The group of homeomorphisms of M^S is G.

Proof: Let h be a homeomorphism of M^S; then, by Lemma 12, h either preserves or reverses the order $\ll$. In the latter case, compose it with the time reflection g defined by

$$g(x_0, x_1, x_2, x_3) = (-x_0, x_1, x_2, x_3),$$

so that $h \circ g$ becomes $\ll$-order-preserving. By Theorem 1, either h or $h \circ g$ belongs to G_0. In any case $h \in G_0 \cup G_0 g^{-1} = G$. This completes the proof.

ACKNOWLEDGMENTS

The author is grateful to Professor R. D. Norman, Professor A. J. Coleman, Professor R. R. D. Kemp, Professor D. Gregery, and Professor N. Rice of Queen's University for their invaluable help and encouragement. Thanks are also due to B. R. Subbarao and R. K. Patnaik for numerous helpful discussions. Financial support from the Canadian Commonwealth Scholarship Committee and Regional Engineering College is gratefully acknowledged.

[1] J. L. Kelley, *General Topology* (Van Nostrand, New York, 1955).
[2] S. Nanda, Ph.D. thesis, Queen's University, Kingston (1969).
[3] E. C. Zeeman, Topology **6**, 161 (1967).
[4] E. C. Zeeman, J. Math. Phys. **5**, 490 (1964).
[5] P. R. Halmos, *Measure Theory* (Van Nostrand, New York, 1950).

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 12 NUMBER 3 MARCH 1971

Gravitational Radiation Damping of Slowly Moving Systems Calculated Using Matched Asymptotic Expansions*

WILLIAM L. BURKE†
California Institute of Technology, Pasadena, California

(Received 13 February 1970)

This paper treats the slow-motion approximation for radiating systems as a problem in singular perturbations. By using the method of matched asymptotic expansions, we can construct approximations valid both in the near zone and the wave zone. The outgoing-wave boundary condition applied to the wave-zone expansion leads, by matching, to a unique and easily calculable radiation resistance in the near zone. The method is developed and illustrated with model problems from mechanics and electromagnetism; these should form a useful and accessible introduction to the method of matched asymptotic expansions. The method is then applied to the general relativistic problem of gravitational radiation from gravitationally bound systems, where a significant part of the radiation can be attributed to nonlinear terms in the expansion of the metric. This analysis shows that the formulas derived from the standard linear approximation remain valid for gravitationally bound systems. In particular, it shows that,according to general relativity,bodies in free-fall motion do indeed radiate. These results do not depend upon *any* definition of gravitational field energy.

I. INTRODUCTION

Problems connected with radiation resistance and radiation damping have had a long and checkered history in classical physics. When early work on the motion of bodies in general relativity was extended by workers hoping to see radiative effects, difficulties appeared and controversies arose, in some cases persisting even until today. Some early workers claimed that there was no radiation damping, others found damping, and a few even found antidamping.

Lemmas 9 and 11 together prove that if T is a timelike or a lightlike line and h a homeomorphism of M^S, then $h \mid T$, i.e., h restricted to T, is either $\ll$-order-preserving or $\ll$-order-reversing. It now remains to be shown that this remains true globally.

Lemma 12: Let T_1 and T_2 be two timelike or (and) lightlike lines and h a homeomorphism of M^S; then T_1 and T_2 are mapped with the same orientation.

Proof: Suppose to the contrary that h preserves the orientation of T_1 and reverses that of T_2. First, assume that T_1 and T_2 intersect at a point x. Take a spacelike hyperplane H at x; then $M^S - H$ has two components, one containing T_1' and T_2' and the other containing T_1'' and T_2'', where $\{x\} \cup T_1' \cup T_1'' = T_1$ and $\{x\} \cup T_2' \cup T_2'' = T_2$. On the other hand, $M^S - hH$ has two components, one containing hT_1'' and hT_2' and the other containing hT_2'' and hT_1'. This is a contradiction. Secondly, if T_1 and T_2 are disjoint, then take a third timelike or lightlike line T_3 which meets both T_1 and T_2. Let $T_1 \cap T_3 = \{x\}$ and $T_2 \cap T_3 = \{y\}$. By applying the argument of the first case at x, one concludes that h preserves the orientation of T_3, and, similarly applying it at y, one concludes that orientation of T_2 is preserved, which is a contradiction. Thus we have proved that the order relation $\ll$ is either preserved or reversed in M.

Theorem 2: The group of homeomorphisms of M^S is G.

Proof: Let h be a homeomorphism of M^S; then, by Lemma 12, h either preserves or reverses the order $\ll$. In the latter case, compose it with the time reflection g defined by

$$g(x_0, x_1, x_2, x_3) = (-x_0, x_1, x_2, x_3),$$

so that $h \circ g$ becomes $\ll$-order-preserving. By Theorem 1, either h or $h \circ g$ belongs to G_0. In any case $h \in G_0 \cup G_0 g^{-1} = G$. This completes the proof.

ACKNOWLEDGMENTS

The author is grateful to Professor R. D. Norman, Professor A. J. Coleman, Professor R. R. D. Kemp, Professor D. Gregery, and Professor N. Rice of Queen's University for their invaluable help and encouragement. Thanks are also due to B. R. Subbarao and R. K. Patnaik for numerous helpful discussions. Financial support from the Canadian Commonwealth Scholarship Committee and Regional Engineering College is gratefully acknowledged.

[1] J. L. Kelley, *General Topology* (Van Nostrand, New York, 1955).
[2] S. Nanda, Ph.D. thesis, Queen's University, Kingston (1969).
[3] E. C. Zeeman, Topology **6**, 161 (1967).
[4] E. C. Zeeman, J. Math. Phys. **5**, 490 (1964).
[5] P. R. Halmos, *Measure Theory* (Van Nostrand, New York, 1950).

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 12 NUMBER 3 MARCH 1971

Gravitational Radiation Damping of Slowly Moving Systems Calculated Using Matched Asymptotic Expansions*

WILLIAM L. BURKE†
California Institute of Technology, Pasadena, California

(Received 13 February 1970)

This paper treats the slow-motion approximation for radiating systems as a problem in singular perturbations. By using the method of matched asymptotic expansions, we can construct approximations valid both in the near zone and the wave zone. The outgoing-wave boundary condition applied to the wave-zone expansion leads, by matching, to a unique and easily calculable radiation resistance in the near zone. The method is developed and illustrated with model problems from mechanics and electromagnetism; these should form a useful and accessible introduction to the method of matched asymptotic expansions. The method is then applied to the general relativistic problem of gravitational radiation from gravitationally bound systems, where a significant part of the radiation can be attributed to nonlinear terms in the expansion of the metric. This analysis shows that the formulas derived from the standard linear approximation remain valid for gravitationally bound systems. In particular, it shows that,according to general relativity,bodies in free-fall motion do indeed radiate. These results do not depend upon *any* definition of gravitational field energy.

I. INTRODUCTION

Problems connected with radiation resistance and radiation damping have had a long and checkered history in classical physics. When early work on the motion of bodies in general relativity was extended by workers hoping to see radiative effects, difficulties appeared and controversies arose, in some cases persisting even until today. Some early workers claimed that there was no radiation damping, others found damping, and a few even found antidamping.

Some claimed that freely falling bodies could not radiate while others saw no difference between free-fall motions and other motions. Attempts to extend the linear theory met logarithmic divergences, and these eroded one's confidence in any of the predictions of the linear theory. Today we understand most of these difficulties, and a consistent, unified picture of gravitational radiation is emerging.

Most early calculations using slow-motion and weak-field approximations were sensitive to the order of approximations, and different results appeared depending upon which limit was taken first. Such paradoxical behavior is typical of a class of perturbation problems called by applied mathematicians "singular perturbations." Paradoxes similar to those found in relativity arose long ago in fluid mechanics and, indeed, provided the motivation for much of modern perturbation theory. One of the major achievements of singular perturbations was the resolution in 1954 by Kaplun of Stokes' paradox, which originated in 1851 and had confused many of the best workers in fluid mechanics.[1] For the slowly moving systems studied here, it will be the limits $r \to \infty$ and $v \to 0$ which require careful treatment.

This paper aims at a straightforward yet complete treatment of the slow-motion limit of radiating systems, by means of singular perturbation theory. The first part of this paper presents the essential background needed to handle "matched asymptotic expansions" and works two simple examples. The rest of the paper applies these methods to more realistic problems taken from classical electromagnetism and general relativity.

Hopefully, the techniques have now been refined to the point where harder but important problems such as slowly-moving but fully relativistic systems and radiation from moving "Schwarzschild" singularities may be tackled. Without the methods presented here, these problems are probably insoluble.

A. Technical Background

For background on the problem of equations of motion in general relativity, the reader should see the review article by Goldberg.[2] For background on gravitational radiation and the problem of coupling it to sources, see Pirani,[3] Bonnor,[4] or Trautman.[5] For background on the slow-motion approximation, see the book by Infeld and Plebanski[6] or the review paper by Einstein and Infeld.[7] Some of the Russian work is discussed in Fock.[8] These all contain many further references.

The slow-motion approximation scheme was first applied in the pioneering paper of Einstein, Infeld, and Hoffman.[9] They studied the motion of point particles, and their approach is now called the EIH method. The slow-motion scheme was applied to fluids by Chandrasekhar,[10] and in that context is called the post-Newtonian approximation. The study of radiation from slowly moving systems has been carried out by many authors, notably Chandrasekhar and his students,[11,12] Trautman,[13] Peres,[14] Peres and Rosen,[15] Carmeli,[16] and Infeld and Michalska-Trautman.[17] While some of the above workers came to the same conclusion as does this paper, they did not recognize that the singular nature of the problem and the manipulations needed to bring in radiation effects are based on ad hoc arguments rather than on routine techniques of singular perturbations. Further, the EIH work depends heavily on the use of "good" δ functions, despite the fact that *nonlinearities* are important in the problem.

To avoid the difficulties involved in treating radiation in the slow-motion limit, some workers have tried to use only a weak-field approximation, which they called the fast-motion approximation. See, in particular, Havas,[18] Havas and Goldberg,[19] and Smith and Havas.[20] Although their method predicted antidamping for the case of gravitationally bound systems, this is not in conflict with the results given here since for weak-field, bound systems the motions *are* slow. Ignoring this fact leads to an incorrect ordering of terms. A crucial part of the radiation damping must have been hidden in the "apparently" higher-order terms of the fast-motion approach. In turn, the work presented here gives little help with the antidamping problem associated with the fast-motion approximation. I can only point out (1) that the fast-motion method is not well suited for gravitationally bound systems (that is, the problem is singular and the weak-field and slow-motion limits do not commute) and (2) that great care is needed in applying coordinate conditions.

My work does not use harmonic coordinates. I find that there are excessively large, time-odd, metric perturbations in the near zone which appear in harmonic coordinates and which can easily be (and were for a time) misinterpreted as antidamping.

Some of the above work on the subject addressed itself to the task of deriving the equations of motion of particles directly from the field equations. I have not attempted to repeat that work here, but I have used the result of such work that particles move along geodesics of the exact metric. I will remark in passing that such derivations were hampered by an apparent restriction of the particles to uniform motion. The resolution of this by allowing *implicit* as well as

explicit dependence on the small parameter can only be justified by passing from power series expansions to the richer field of asymptotic expansions. Indeed, such tricks are common practice in singular perturbation theory.

Expansions using the two small parameters of weakness and slowness were used also by Bonnor and others[21,22] in what they called the "double series" method. Unfortunately, their work did not cover gravitationally bound systems. For those systems there is a definite relationship between the two parameters, and hence we have only one independent small parameter.

B. Summary

The primary result of the work presented here is a method for treating radiation resistance in a slow-motion approximation. The method is applied here to the problem of gravitational radiation produced by systems moving under the influence of purely gravitational forces. For such systems, just as for weak-field systems dominated by nongravitational forces (and hence described by linearized gravity), the escape of radiation is accompanied by the presence of resistive forces which react on the radiating system and extract the energy that appears in the radiation. This energy loss is related to the changing quadrupole moment by the same formula as that which is derived using the linearized theory.

Our work differs from the work cited above mainly in its use of singular perturbation techniques such as matched asymptotic expansions to put the approximation scheme on a systematic foundation. This approach emphasizes the role of the nonlinearities in providing a replacement for the mechanical stresses absent in "free-fall" motion. It also emphasizes the care needed in choosing appropriate coordinate conditions (gauge conditions).

Further advantages of this method (although not unique to it) are (1) the ability to compute the damping terms without having to compute the more complicated post-Newtonian and post-post-Newtonian terms, (2) the use of resistive forces as opposed to energy conservation to compute the damping, (3) the easy avoidance of singular functions, which have no place in nonlinear problems, and (4) the completeness of the final solution, which is a uniformly valid approximation both in the near zone and in the wave zone.

The introductory section of this paper on singular perturbations (Sec. II) provides a simple and readily accessible introduction to the ideas of matched asymptotic expansions. The discussion of the spurious solutions introduced by the approximation scheme should also be of interest to workers in singular perturbations and to those worried by the runaway solutions of classical electromagnetism. Tensor spherical harmonics are used to simplify the calculations, and a brief appendix on these is included.

II. THE SLOW-MOTION EXPANSION AS A SINGULAR PERTURBATION PROBLEM

The method of matched asymptotic expansions, which we will use to study radiation resistance, is applicable to electromagnetic problems, acoustic problems, and a great variety of other wave-propagation problems. This section will describe the method and present two simple examples illustrating the ideas. The book by Cole[23] should be consulted for a more complete treatment of the subject.

A. Asymptotic Expansions; Regular and Singular Perturbation Problems

It is very important to realize that the expansions we will use to find approximate solutions to our equations are asymptotic expansions rather than convergent power-series expansions. Our techniques for matching expansions together will rely on the interpretation of the expansions as asymptotic expansions.

To define the terminology, consider a function $f(x, \epsilon)$ of one or several coordinates denoted by x and of a small parameter ϵ. An asymptotic expansion of $f(x, \epsilon)$,

$$f(x, \epsilon) \sim \sum_{k=0}^{\infty} f_k(x, \epsilon), \tag{1}$$

is an expansion having the property that

$$\lim_{\epsilon \downarrow 0} \left(f(x, \epsilon) - \sum_{k=0}^{m} f_k(x, \epsilon) \right) \Big/ \gamma_m(\epsilon) = 0, \tag{2}$$

where the $\gamma_m(\epsilon)$ are called "gauge functions." They must satisfy

$$\gamma_{n+1}(\epsilon) \ll \gamma_n(\epsilon), \tag{3}$$

and they provide a set of standard orders of magnitude. Here we will be able to use the powers of ϵ as gauge functions, but functions such as $\epsilon \log \epsilon \cdots$ occur frequently in practice. The infinite sum in Eq. (1) is a formal sum which may or may not converge. Convergence is replaced by the more general property (2). From a practical point of view, a slowly convergent expansion is much less useful than a divergent expansion whose first few terms nevertheless give a numerically good approximation for the range of ϵ of interest. Such a divergent expansion is asymptotic in the sense of Eq. (2).

The simplest form of the f_k is, of course,

$$f_k(x, \epsilon) = \epsilon^k f_k(x). \tag{4}$$

However, an expansion of type (1) is of interest only if the limit in (2) is uniform over the range of x considered. If the assumption (4) gives such a uniformly valid expansion, we call the perturbation problem regular; otherwise, we call it singular. The usual slow-motion expansion for radiating systems is singular in that the limit (2) is not uniform for large radii.

B. A Trivial Singular Expansion; Matching

In a singular problem an expansion of type (4) is not uniformly valid typically in regions such as $x = O(\epsilon)$ or $x = O(1/\epsilon)$. In such cases a different expansion may be used to represent the function in the regions of nonuniformity. These multiple expansions may be combined into a composite expansion, uniformly valid in the whole x domain considered, which does not have the simple form (4) but which is still a special case of (1). However, for our purpose it will be sufficient to use the multiple expansions without constructing the corresponding composite expansion explicitly.

Nonuniformities come up in the expansions of even very simple functions. Suppose we want the asymptotic expansion for small positive ϵ of the function[24]

$$f(x, \epsilon) = 1 + x + \epsilon/x. \tag{5}$$

If one assumes an expansion in the sequence

$$f(x, \epsilon) \sim f_0(x) + \epsilon f_1(x) + \cdots, \tag{6}$$

one can find the coefficients $f_i(x)$ by repeated use of the limit property (2). The leading term is

$$f_0(x) = 1 + x. \tag{7}$$

We can see from Fig. 1 that f_0 is a very poor approximation near $x = 0$. This could have been anticipated because the limit used to evaluate f_0 was not uniform as $x \to 0$. We call expansion (6) the outer expansion, and the limit $\epsilon \to 0$ for fixed x the outer limit.

We could take the limit $\epsilon \to 0$ in such a manner that x also goes to zero at the same rate. This results in a different ordering of terms and a different expansion. This limit, called the inner limit, is most easily taken by rewriting our function in terms of a new variable

$$X \equiv x/\epsilon. \tag{8}$$

The inner limit is then $\epsilon \to 0$ for fixed X.

Assuming an inner expansion

$$f(\epsilon X, \epsilon) = F(X, \epsilon) \sim F_0(X) + \epsilon F_1(X) + \cdots, \tag{9}$$

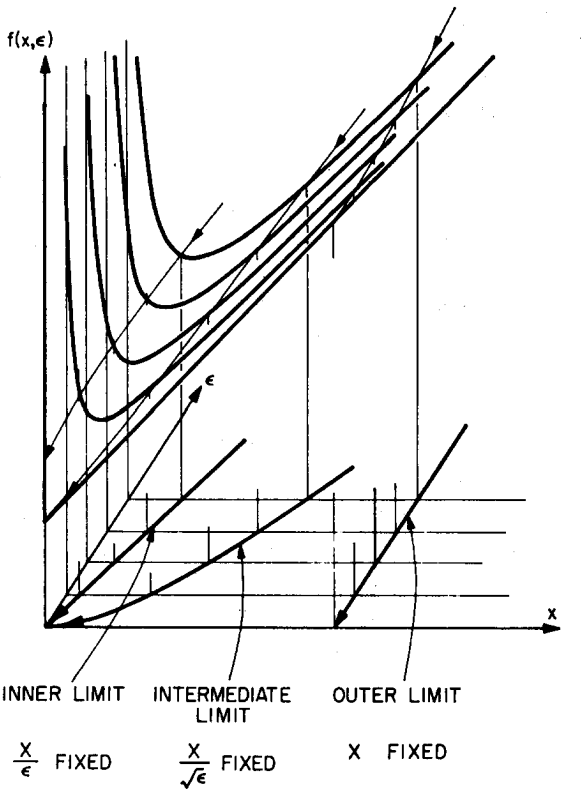

FIG. 1. A plot of $f(x, \epsilon) = 1 + x + \epsilon/x$, showing the various limit processes described in the text. Notice that the figure is 3-dimensional with axes x, ϵ, and f.

one finds for the leading term

$$F_0(X) = 1 + 1/X. \tag{10}$$

Again, the limit is not uniform; the inner expansion is not uniformly valid for $X = O(1/\epsilon)$.

Singular behavior quite similar to this will occur in the expansions that we will make for the fields produced by slowly moving sources. There we will need to carry the wave-zone boundary conditions from an outer expansion into an inner expansion. This will be done by *matching*, one of the most important concepts in singular perturbations.

The basic idea of matching is that there should be an intermediate region where both the inner and the outer expansions are valid approximations. If they both are close to the exact solution in this region, then they must also be close to each other. For our model function this intermediate region must be one where x is small but X is large. Looking at our inner and outer solutions [Eqs. (7) and (10)] and the behavior of f shown in Fig. 1, we see that a plausible relation is

$$F_0(\infty) = f_0(0). \tag{11}$$

This relation is only a special case of a more general matching principle.

Suppose we look at a function $\eta(\epsilon)$ intermediate between 1 and ϵ,

$$\epsilon \ll \eta(\epsilon) \ll 1, \tag{12}$$

for example $\sqrt{\epsilon}$, and define intermediate variables

$$x_\eta \equiv x/\eta(\epsilon). \tag{13}$$

For our intermediate region we pick the region where x_η is of order unity; there x is of order $\eta(\epsilon)$ (hence small as $\epsilon \to 0$), and X is of order $\eta(\epsilon)/\epsilon$ (hence large as $\epsilon \to 0$). We define an intermediate limit

$$\lim_\eta \equiv \text{limit } \epsilon \downarrow 0 \text{ for } x_\eta \text{ fixed}, \tag{14}$$

and for a matching condition require that the solutions be close in the intermediate limit, that is

$$\lim_\eta (F_0 - f_0) = 0. \tag{15}$$

This matching condition coincides with our intuitive guess, Eq. (11). A typical intermediate limit is indicated in Fig. 1. An extensive discussion of the heuristic ideas underlying such formal matching can be found in the work of Kaplun.[25]

C. A Nontrivial Example of Matching

We can illustrate the ideas of matching and show their applicability to radiation problems by considering a highly idealized model problem. Here we will solve for the damping of a mechanical oscillator coupled to an infinite elastic string, the oscillator losing energy by radiating waves out along the string. The geometry of the situation is sketched in Fig. 2.

For small slopes the transverse displacement of the string obeys the equation

$$\rho \frac{\partial^2 y}{\partial t^2} - T \frac{\partial^2 y}{\partial x^2} = f(x, t) \tag{16}$$

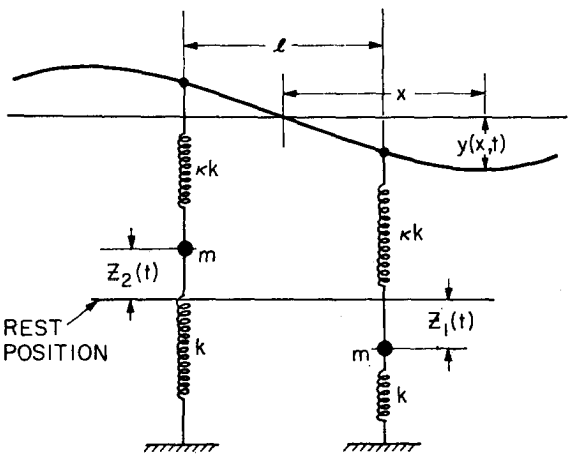

FIG. 2. The geometry of the mass-spring oscillators coupled to an elastic string used as a model radiating system.

where

y = transverse displacement of the string,
x = longitudinal coordinate along the string,
ρ = mass per unit length, assumed constant,
T = tension, assumed constant,
f = transverse force arising from the coupling spring.

Two identical spring-mass oscillators are weakly coupled to the string and are separated by a distance l. We need to consider two oscillators in order to introduce a length scale for the source into the problem. This length together with the wavelength of the waves on the string enables us to form a useful dimensionless parameter.

The relevant dimensioned parameters in this problem are the wave speed on the elastic string, a, given by

$$a \equiv (T/\rho)^{\frac{1}{2}}, \tag{17}$$

and the natural frequency of the oscillators, ω, given in terms of the mass m and spring constant k by

$$\omega \equiv (k/m)^{\frac{1}{2}}. \tag{18}$$

We can combine a and ω to form a length λ, the wavelength of waves on the string generated by a source moving with frequency ω:

$$\lambda \equiv 2\pi a/\omega. \tag{19}$$

We shall now restrict our attention to systems where $l \ll \lambda$ and use their ratio

$$\epsilon \equiv l/\lambda \tag{20}$$

as our expansion parameter.

The other important dimensionless parameters in the problem are κ, the ratio of the stiffness of the coupling springs to the stiffness of the oscillator springs (cf. Fig. 2), and σ, the ratio of the stiffness of the coupling springs to the stiffness of the string:

$$\sigma \equiv \kappa k l/2T = 2\pi^2 \kappa \epsilon^2 m/\rho l. \tag{21}$$

To simplify the calculations, we shall assume that the parameters κ and σ are both much smaller than the parameter ϵ. Physically this means that the masses and the string are very weakly coupled to each other.

To find the motion of the system, we proceed as follows: (i) For any given motion of the masses, we compute the motion of the string induced by its weak coupling to the masses; (ii) then we find the force exerted by the string on the masses and write an equation of motion for the masses. This equation of motion will contain damping terms induced by the radiation on the string, but will not contain the string coordinate $y(x, t)$ explicitly.

Let us concentrate on the odd mode, having the property that

$$Z(t) \equiv Z_1(t) = -Z_2(t), \tag{22}$$

where $Z_1(t)$ and $Z_2(t)$ are the transverse positions of the two masses relative to their equilibrium positions. We can do this most easily by imposing the boundary condition

$$y(0, t) = 0 \tag{23}$$

and restricting our attention to positive values of x. The forcing term in Eq. (16) is

$$f(x, t) = \kappa k[Z - y(l/2, t)]\delta(x - l/2). \tag{24}$$

Using the physical parameters of the system, we can transform to dimensionless coordinates that are natural for the system. We introduce a dimensionless time

$$t^* \equiv \omega t \tag{25}$$

and two dimensionless position coordinates

$$x^* \equiv 2\pi x/\lambda, \quad X^* \equiv 2\pi x/l = x^*/\epsilon. \tag{26}$$

Since X^* is normalized by means of l, it is a suitable variable for describing the string near the oscillators; it is an *inner* variable. On the other hand, x^* is defined in terms of the properties of the radiation; it is an *outer* variable, suitable for $x \gg l$. The fact that different physical processes dominate in different regions leads to the use of two different distance variables; correspondingly, we also expect that we shall need two different asymptotic expansions to describe the motion of the string.

In outer (x^*, t^*) coordinates the equation of motion for the string is

$$\frac{\partial^2 y}{\partial t^{*2}} - \frac{\partial^2 y}{\partial x^{*2}} = 0. \tag{27}$$

Note that we expect the outer expansion to be valid only for $x^* \gg \epsilon$. This region does not include the oscillator location $x^* = \epsilon\pi$; hence there is no forcing term in Eq. (27). The outer solution knows about the oscillator only through matching with the inner expansion.

In inner (X^*, t^*) coordinates the wave equation for the string reads

$$\frac{\partial^2 y}{\partial X^{*2}} - \epsilon^2 \frac{\partial^2 y}{\partial t^{*2}} = -\frac{\sigma}{\pi}(Z - \zeta)\delta(X^* - \pi), \tag{28}$$

where we have introduced the function

$$\zeta(t^*) \equiv y(l/2, t^*/\omega) \tag{29}$$

to represent the displacement of the string at the oscillator location.

Note that since we deal with linear equations, we may let Z, y, and ζ be dimensional and not worry about their magnitude relative to l or λ.

The mechanical equation describing the motion of the mass is

$$\frac{d^2 Z}{dt^{*2}} + Z + \kappa(Z - \zeta) = 0. \tag{30}$$

We proceed to find an approximation to the string displacement for an arbitrary motion of the mass. We assume that there is an asymptotic expansion for y valid in the inner limit:

$$y \sim A(X^*, t^*) + \epsilon B + \epsilon^2 C \cdots. \tag{31}$$

Inserting this expansion into the inner equation (28) and using the limit process to collect terms of the same order, we obtain the following equations for the terms in the inner expansion:

$$\frac{\partial^2 A}{\partial X^{*2}} = -\frac{\sigma}{\pi}(Z - \zeta)\delta(X^* - \pi), \tag{32a}$$

$$\frac{\partial^2 B}{\partial X^{*2}} = 0, \tag{32b}$$

$$\frac{\partial^2 C}{\partial X^{*2}} = \frac{\partial^2 A}{\partial t^{*2}}. \tag{32c}$$

Looking at the leading equation, we can see why this is sometimes called the "quasistatic" limit; time enters this equation only as a parameter.

The solution for A is

$$A = \begin{cases} \sigma(Z - \zeta) & \text{for} \quad X^* \geq \pi \\ (\sigma/\pi)X^*(Z - \zeta) & \text{for} \quad 0 \leq X^* \leq \pi \end{cases}. \tag{33}$$

Since $\sigma \ll \epsilon$ and $A = O(\zeta)$, it follows that we can neglect ζ relative to Z in (33) and write

$$A \sim \begin{cases} \sigma Z & \text{for} \quad X^* \geq \pi \\ \sigma Z X^*/\pi & \text{for} \quad 0 \leq X^* \leq \pi \end{cases}. \tag{34}$$

In particular, a first approximation for the string displacement at the point of coupling is

$$\zeta(t^*) \sim \sigma Z(t^*). \tag{35}$$

We have partially solved our problem. Given any motion of the masses, $Z(t^*)$, we have found the first term in the inner expansion for the motion of the string. Since this term is time symmetric, there is no information about damping in the expansion so far.

Using the boundary condition at $X^* = 0$, we see that the solution B to Eq. (32b) must have the form

$$B = \alpha(t^*)X^*. \tag{36}$$

One is tempted to choose $\alpha = 0$ to avoid a divergent behavior for large X^*. On the other hand, one should

expect on the basis of our previous discussion that the inner expansion might not be uniformly valid in the limit $X^* \to \infty$, since that limit takes us into the outer region where other phenomena dominate the equation of motion. If the approximation is not uniformly valid for large distances, then this divergent behavior does not mean that the solution actually diverges. One can only determine whether or not such a divergent term is present by examining the problem in the outer limit and then matching the outer expansion with the inner expansion.[26]

We assume that there is an asymptotic expansion valid in the outer limit:

$$y \sim F(x^*, t^*) + \epsilon G + \cdots . \tag{37}$$

All the terms of this expansion must satisfy equations of the form

$$\frac{\partial^2 F}{\partial x^{*2}} - \frac{\partial^2 F}{\partial t^{*2}} = 0. \tag{38}$$

We can write the solutions to this equation as traveling waves

$$F(x^*, t^*) = W(t^* \mp x^*). \tag{39}$$

The function W will be determined by matching; the upper sign corresponds to outgoing waves at infinity, and the lower sign corresponds to incoming waves.

To match this first term of the outer expansion with the inner solution, we first expand it for small x^*,

$$y_{\text{out}} \to W(t^*) \mp x^* W'(t^*) + \cdots , \tag{40}$$

and then rewrite it in inner coordinates,

$$y_{\text{out}} \to W(t^*) \mp \epsilon X^* W'(t^*) + \cdots . \tag{41}$$

Expanding the inner solution, Eqs. (34) and (36), for large X^*, we find

$$y_{\text{in}} \to \sigma Z(t^*) + \epsilon \alpha(t^*) X^* + \cdots . \tag{42}$$

The function W appearing in the outer expansion is determined by matching the zero-order terms to find

$$W(t^*) = \sigma Z(t^*), \tag{43}$$

and the function $\alpha(t^*)$ appearing in the first-order term of the inner expansion is then determined by matching the first-order terms to find

$$\alpha(t^*) = \mp \sigma \frac{dZ}{dt^*}. \tag{44}$$

Thus we see that actually the function B of Eq. (36) is nonzero; it is determined uniquely by the wave-zone boundary condition.

From these results we can find a more accurate expression for $\zeta(t^*)$ than that given by Eq. (35):

$$\zeta(t^*) \sim \sigma Z(t^*) \mp \epsilon \pi \sigma \frac{dZ}{dt^*}. \tag{45}$$

The $O(\epsilon)$ correction to $\zeta(t^*)$ is time odd, i.e., it changes sign when the boundary condition is changed from outgoing waves to incoming waves. Thus this term includes the effects of the irreversible loss of energy in the radiation.

Inserting this expression for $\zeta(t^*)$ into our force law (30), we find a more accurate equation of motion for the oscillator:

$$\frac{d^2 Z}{dt^{*2}} + \epsilon \kappa \sigma \pi \frac{dZ}{dt^*} + (1 + \kappa) Z = 0, \tag{46}$$

(taking the upper sign corresponding to the physically interesting, outgoing-wave boundary condition). The damping causes the motion to decay, with a Q (Q = number of radians to lose $1/e$ of the energy) given by

$$Q \approx (\pi \epsilon \kappa \sigma)^{-1} = 2T\lambda/\pi \kappa^2 k l^2. \tag{47}$$

If one carries this procedure on to higher orders, one discovers that all of the higher terms in the outer expansion vanish and that a more accurate expression for $\zeta(t^*)$ is

$$\zeta(t^*) \sim \sigma(Z \mp \epsilon \pi Z' + \tfrac{1}{2}\epsilon^2 \pi^2 Z'' \mp \tfrac{1}{6}\epsilon^3 \pi^3 Z''' + \cdots). \tag{48}$$

This expression leads to a more accurate equation of motion (by use of the upper sign):

$$(\tfrac{1}{6}\epsilon^3 \pi^3 \sigma \kappa) Z''' + (1 + \tfrac{1}{2}\epsilon^2 \pi^2 \kappa \sigma) Z'' + \epsilon \kappa \sigma \pi Z' + (1 + \kappa) Z = O(\epsilon^4). \tag{49}$$

This is a third-order equation, whose solutions consist of two weakly damped, oscillatory modes which are perturbations of the modes of the unperturbed oscillators, plus a third mode which behaves approximately like

$$Z \propto \exp(-6t^*/\pi^3 \epsilon^3 \kappa \sigma). \tag{50}$$

As the equation of motion is made more accurate by including more and more terms, (i) more and more modes of this third type appear, and (ii) each additional term in the equation of motion produces a correction to modes of this type which is larger than the last correction. The net effect of these corrections is to make the complex frequencies of these additional modes trace out diverging spirals in the complex frequency plane. By contrast, the complex frequencies corresponding to the two original modes trace out converging spirals. This leads us to suspect that the

modes introduced by the higher derivatives are spurious artifacts of the approximation. We can verify this suspicion by looking at solutions valid for arbitrary ϵ.

Our problem can be solved without making the assumption that $\epsilon \ll 1$. Treating only κ and σ as small, one can derive an equation of motion

$$\frac{d^2Z}{dt^{*2}} + (1+\kappa)Z - \frac{\kappa\sigma}{2\pi\epsilon} \times \left(\int^{t^*} Z(\tau)\,d\tau - \int^{t^*} Z(\tau - 2\pi\epsilon)\,d\tau\right) = 0, \quad (51)$$

involving a time-delayed term. Equations of this type are called delay-differential equations.[27] Exponential solutions

$$Z = \exp(pt^*) \quad (52)$$

to this equation satisfy the characteristic equation

$$p^2 + (1+\kappa) - (\kappa\sigma/2\pi\epsilon p)(1 - e^{-2\pi\epsilon p}) = 0, \quad (53)$$

which has infinitely many roots. In addition to the roots which are perturbations of the free oscillator frequencies, $p = \pm i$, there are infinitely many roots with large damping, one of which is approximately

$$p \approx -(2\pi\epsilon/9)\log(2\pi\epsilon/\kappa\sigma), \quad (54)$$

with a logarithmic ϵ-dependence. These roots correspond to modes having an integral number of wavelengths trapped between the poorly reflecting points where the oscillators are attached, and leaking away rapidly. These modes clearly violate our assumption $l \ll \lambda$, and the attempts of our approximation scheme to describe these modes lead to the spurious solutions, which are meaningless. Compare Eq. (50) with Eq. (54).

We can now understand the nature of the approximate equations (46) and (49), which result when the integral in Eq. (51) is replaced by derivatives (through expansion in powers of ϵ). These equations generate corrections to the free-oscillator modes that are more accurate as more terms are kept. At the same time, new, spurious modes appear which violate the assumptions under which solutions of the approximate equations are valid approximations to solutions of the exact equations.

Thus we can use approximate equations such as Eqs. (46) and (49) as long as we ignore[28] the spurious modes introduced by our approximation scheme [which we can see result from rashly expanding the exponential in Eq. (53) even when its argument is not small]. Caution is required in using the approximate equations since some of the spurious modes that appear grow exponentially (e.g., the infamous runaway solutions of electromagnetism) and would lead to instabilities if the equations were being integrated numerically.[29]

From this example we see that the solution to a problem having length scales of different magnitudes may be approximated by an asymptotic expansion in the dimensionless ratio of these lengths. However, the existence of two distinct characteristic length scales and thus of two different dimensionless coordinates is a warning that the problem may be singular and may require the techniques of matching.[30]

III. SLOW-MOTION ELECTROMAGNETISM

We can simplify the application of matched asymptotic expansions to gravity by first discussing its application to electromagnetism, where the ideas and language are familiar. The gravity section can then concentrate on those aspects of the problem peculiar to gravity. The details of slow-motion electromagnetism using matching form a useful exercise, or they may be found in the author's PhD thesis.[31] Here we shall give only an outline of the analysis.

A. The Slow-Motion Formalism

We can get a system of equations very similar to our stretched-string equation if we represent the electromagnetic field by a vector potential in Lorentz gauge. Thus we shall work with the field equations ($c = 1$)

$$A^{\mu}{}_{;\nu}{}^{\nu} = -4\pi j^{\mu}, \quad (55)$$

the gauge conditions

$$A^{\nu}{}_{;\nu} = 0, \quad (56)$$

and the force law

$$\rho a^{\mu} = -g^{\mu\nu}(A_{\nu;\sigma} - A_{\sigma;\nu})j^{\sigma}, \quad (57)$$

where

j^{σ} = 4-current density,

ρa^{μ} = 4-force density.

In the slow-motion limit, the space and time components of the vector potential are of different orders. We shall keep the sizes of terms explicit by using the familiar decomposition into 3-vectors plus scalars.

We restrict our attention to sources of the electromagnetic field which move slowly. Here, as before, our system has two distinct length scales, the length l, characteristic of the size of the source, and the length $\lambda = 2\pi c/\omega$, characteristic of the motion of the source and the resulting radiation. Our slowly moving systems satisfy $l \ll \lambda$.

The separation of space into two regions is familiar to students of electromagnetic theory, who refer to the inner region as the induction zone and the outer region as the wave zone. The equations for the terms

in the inner expansion will be similar to the inner equations for the string, Eq. (32), and will start off with the Poisson equation of electrostatics. The equations for the terms in the outer expansion will be homogeneous wave equations.

Several points should be mentioned about the slow-motion formalism used here. The slow-motion expansions appearing in the literature are often written as expansions in inverse powers of the speed of light. This is an unfortunate practice. One usually expands in a dimensionless parameter if possible. An expansion using a dimensioned parameter usually turns out to be a coordinate expansion, which cannot be made uniformly valid over the entire range of coordinates. Here we are definitely dealing with an expansion in a dimensionless parameter.

Another source of confusion arises from using "$1/c$" as an expansion parameter when an outer expansion is necessary. This outer expansion is an expansion in the same small parameter as appears in the inner expansion; yet in the outer limit the $1/c$ terms are not considered small compared with the other terms. Further, ϵ has a definite value, whereas it is more convenient to work in units where $c = 1$. Therefore, I have abandoned the historical practice (EIH, Chandrasekhar, etc.) of writing the expansion as one in the parameter "$1/c$."

Another point concerns our treatment of vectors. One steeped in the spirit of relativity might be tempted to transform the vector components (i.e., change from basis vectors $\partial/\partial X^*$ to $\partial/\partial x^*$) when going between inner and outer coordinates. If we were to do this, the magnitudes of the components of vectors would not correspond to their actual physical magnitudes, and, in addition, we would have to introduce a great deal of useless dual notation such as inner and outer vector spherical harmonics. Instead, we have chosen to refer vectors to basis vectors with unit physical length, i.e., the basis vectors determined by the physical coordinates r and t.

B. Electric Dipole Radiation

Let us look briefly at the slow-motion analysis for a system which emits electric dipole radiation. In the slow-motion limit,electric dipole radiation will dominate the other multipoles in the radiation zone. Specifically, let us look at the fields and forces resulting from the slow motion of a "point" charge along the z axis.

Let the position of the charge be given by $Z(t)$. Then the charge and current distributions are given by

$$\rho = (q/l^3)\delta(X^*)\delta(Y^*)\delta(Z^* - U), \tag{58}$$

$$\mathbf{J} = (\epsilon Q/l^3)\mathbf{e}_z U'(t^*)\delta(X^*)\delta(Y^*)\delta(Z^* - U). \tag{59}$$

Here we have introduced dimensionless inner coordinates

$$t^* \equiv t/\lambda, \quad X^* \equiv x/l \text{ etc.}, \quad U(t^*) = Z(\lambda t^*)/l, \tag{60}$$

and we have used a prime for the derivative of a function with respect to its argument.

Assuming an inner expansion for the scalar potential,

$$\varphi \sim a + \epsilon b + \cdots, \tag{61}$$

we solve Poisson's equation

$$\nabla^2 a = -4\pi\rho \tag{62}$$

to find the leading term

$$a = q[l(R^{*2} + U^2 - 2UR^* \cos\theta)]^{-\frac{1}{2}}. \tag{63}$$

Here R^* is a dimensionless inner coordinate,

$$R^* \equiv r/l. \tag{64}$$

For the vector potential, we assume an inner expansion

$$\mathbf{A} \sim \epsilon\mathbf{M} + \epsilon^2\mathbf{N} + \cdots, \tag{65}$$

which starts with a term of order ϵ since $\mathbf{J}$ is of order $\epsilon\rho$. This keeps a and $\mathbf{M}$ the same size. Solving the near-zone equation for the lowest-order term in the vector potential,

$$\nabla^2\mathbf{M} = -4\pi\mathbf{J}, \tag{66}$$

we find

$$\mathbf{M} = qU'\mathbf{e}_z[l(R^{*2} + U^2 - 2R^*U\cos\theta)]^{-\frac{1}{2}}. \tag{67}$$

For matching we will need the behavior of these solutions for large R^*. Expanding, we find

$$a \to q/lR^* + qU(t^*)\cos\theta/l^2R^{*2} + \cdots, \tag{68}$$

$$\mathbf{M} \to qU'(t^*)\mathbf{e}_z/lR^*. \tag{69}$$

The static, monopole part of this expansion matches to a static, monopole outer solution with no difficulty. The time-dependent dipole terms will radiate, and we will need an outer solution corresponding to electric-dipole radiation. Note an advantage of using two different expansions: The multipole decomposition need only be made in the outer zone, and only there do the low-order multipoles dominate.

Using spherical harmonics,[32] we can write down the potentials representing electric-dipole radiation directly. The scalar potential must be proportional to Y_{10} (axisymmetry implies that $M = 0$). The vector potential must be a combination of $\mathbf{Y}_{100}$ and $\mathbf{Y}_{120}$. ($\mathbf{Y}_{110}$ has the wrong parity and belongs to the magnetic-dipole solution.) The outer limit of the inner expansion of the vector potential $\mathbf{M}$, Eq. (69), has the angular dependence $\mathbf{Y}_{100}$ [recall that $\mathbf{Y}_{100} = e_z/(4\pi)^{\frac{1}{2}}$]; thus the outer solution that matches to it must be

proportional to $\mathbf{Y}_{100}$. The radial dependence of the outer solution is determined by the wave equations and the boundary conditions at $r^* = \infty$. (Wave equations and their multipole solutions are discussed in the Appendix.) The coefficient relating the size of the vector potential to the size of the scalar potential can be determined from the gauge condition.

Assuming outer expansions

$$\varphi \sim P + \cdots, \tag{70}$$

$$\mathbf{A} \sim \mathbf{P} + \cdots, \tag{71}$$

we have for the electric-dipole part of the outer solution,

$$P = Y_{10}[\pm W'(t^* \mp r^*)/r^* + W(t^* \mp r^*)/r^{*2}], \tag{72}$$

$$\mathbf{P} = (\sqrt{3})\mathbf{Y}_{100}W'(t^* \mp r^*)/r^*. \tag{73}$$

The function W will be determined by matching, the upper sign corresponds to outgoing waves at infinity, and r^* is a dimensionless radial coordinate,

$$r^* \equiv r/\lambda. \tag{74}$$

By matching the small-r^* behavior of the outer solution, Eqs. (72) and (73), to the large-R^* behavior of the inner solution, Eqs. (68) and (69), we determine $W(t^*)$:

$$W(t^*) = \epsilon^2 q(\tfrac{4}{3}\pi)^{\frac{1}{3}}U(t^*)/l. \tag{75}$$

With this solution for $W(t^*)$, Eqs. (72) and (73) describe the radiation produced by a given motion $U(t^*)$ of the charge.

Expanding the outer solution, Eqs. (72) and (73), for small r^* to higher orders, we finally come to terms that are sensitive to the outgoing-wave boundary condition; these are the first terms that can lead to radiation damping:

$$P \to \cdots \pm \epsilon^2 q \cos\theta r^* U'''(t^*)/3l + \cdots, \tag{76}$$

$$\mathbf{P} \to \cdots \mp \epsilon^2 q \mathbf{e}_z U''(t^*)/l + \cdots. \tag{77}$$

As in the string problem, these time-odd terms match onto terms satisfying homogeneous equations in the inner zone. The leading time-odd terms in the inner expansion are thus uniquely determined by the homogeneous equations

$$\nabla^2 \varphi_R = 0, \tag{78}$$

$$\nabla^2 \mathbf{A}_R = 0 \tag{79}$$

and the matching requirements

$$\varphi_R \to \epsilon^3 q R^* \cos\theta U'''(t^*)/3l, \tag{80}$$

$$\mathbf{A}_R \to \epsilon^2 q \mathbf{e}_z U''(t^*)/l. \tag{81}$$

I refer to the time-odd potentials in the inner expansion determined by these equations as "resistive potentials"; for the case of outgoing waves, they are given by

$$\varphi_R = \epsilon^3 q R^* \cos\theta U'''(t^*)/3l, \tag{82}$$

$$\mathbf{A}_R = \epsilon^2 q \mathbf{e}_z U''(t^*)/l. \tag{83}$$

The force on the charge due to the resistive fields coming from these potentials is

$$\mathbf{F}_R = \tfrac{2}{3}q^2 \mathbf{e}_z \frac{d^3\mathbf{Z}}{dt^3} + \cdots. \tag{84}$$

The scale lengths l and λ have dropped out of this expression for the force on the charge caused by the escape of radiation. They were used only to achieve the correct ordering of terms.

We can use the resistive field given in Eq. (84) to write an equation of motion for the charge in which all of the degrees of freedom of the electromagnetic field have been eliminated. For example, attaching our charge to a spring gives us a charged mechanical oscillator whose motion will be described by the equation

$$m\frac{d^2Z}{dt^2} + kZ - \tfrac{2}{3}q^2\frac{d^3Z}{dt^3} + \cdots = 0, \tag{85}$$

and from this we could compute the motion of the oscillator,

$$Z(t) = Ae^{i\omega t}\exp(-2\pi r_e \omega t/3\lambda), \tag{86}$$

having a Q

$$Q \approx 3\lambda/4\pi r_e, \tag{87}$$

where

$$r_e \equiv q^2/2m. \tag{88}$$

The approximate equation of motion (85) will have a number of spurious solutions similar to those found for the string equation. All modes which are not perturbations of the zero-order, undamped, near-zone modes should be ignored. A careful check shows that they grow on a time scale so short as to violate the requirement $l \ll c/\omega$. Again these arise because the problem has solutions which do not satisfy our slow-motion requirement and which our slow-motion approximation therefore cannot handle correctly.

One can simplify the calculation of the resistive forces by making use of the invariance of the result under a gauge transformation:

$$\mathbf{A} \to \tilde{\mathbf{A}} = \mathbf{A} + \nabla\chi, \tag{89}$$

$$\varphi \to \tilde{\varphi} = \varphi - \frac{\partial\chi}{\partial t}. \tag{90}$$

It is easy to find a χ such that the new potentials in the outer region have the form

$$\tilde{\varphi} = \alpha\varphi, \tag{91}$$

$$\tilde{\mathbf{A}} = \tilde{W}(r, t)\mathbf{Y}_{L,L+1,M} \tag{92}$$

for any electric multipole. The resistive forces can then be computed solely from $\tilde{\varphi}$, since the vector potential $\tilde{\mathbf{A}}$ has a radial dependence corresponding to spherical harmonics of index $L + 1$, rather than to $L - 1$ as it did before the gauge transformation; its forces in the induction zone are thus smaller than the forces due to the scalar potential by a factor of ϵ^2.

A gauge transformation similar to this will be used in the gravity case to greatly simplify both the calculation and the interpretation. Actually, the above gauge transformation introduces some time-symmetric terms in the near-zone expansion of the vector potential that are $O(1/\epsilon)$. These do not bother us where we deal with linear equations or in very-weak-field gravity, or they can be avoided by only transforming the time-odd part of the field.

A great deal can be learned from the similarity between gravity and electromagnetism, and even closer analogies than those discussed in this paper result from the study of electric quadrupole systems, such as systems where q/m is the same for all particles. Such systems are discussed in the author's PhD thesis,[31] and are recommended to the reader as an instructive exercise.

IV. SLOW-MOTION GRAVITY

A. The Formalism of Slow-Motion Gravity

Unlike our previous examples, Einstein's theory of gravitation[33] is nonlinear. This means that the absolute strength of the field is now important. This field strength is characterized by a second dimensionless parameter κ, related to the typical size of the Newtonian potential U by

$$\kappa \equiv U/c^2. \tag{93}$$

The nature of our approximation scheme depends on the relative sizes of the slowness parameter ϵ and the weakness parameter κ.[34]

One type of expansion is appropriate for systems having very weak fields, that is, where

$$\kappa \ll \epsilon^2. \tag{94}$$

Nongravitational forces dominate such systems, and their behavior in lowest order is described by the linearized theory of gravity.

Another type of system has fields that are not so weak, but instead satisfy

$$\kappa = O(\epsilon^2). \tag{95}$$

Gravitational forces can play a dominant role in such systems, and a description of their behavior requires the inclusion of some nonlinearities even when computing the lowest-order gravitational radiation.

One can also conceive of systems for which the field strengths are $O(1)$, i.e., fully nonlinear, and yet whose motions are still slow. However, these are very special systems, and, moreover, the approximate equations describing such systems are still nonlinear.

Here we confine our attention to systems for which

$$\kappa = O(\epsilon^2). \tag{96}$$

We will see that there is no change in the damping results as one passes from very-weak-field systems to gravitationally bound systems which satisfy Eq. (96).

We assume that the exact metric ${}^*g_{\mu\nu}$ can be expanded as a perturbation about the background metric $g_{\mu\nu}$, which here we take to be flat space (represented in polar coordinates). We take the expansions

$${}^*g_{tt} \sim -1 + \tfrac{1}{2}\epsilon^2\psi + \epsilon^4(\tfrac{1}{2}b + \tfrac{1}{2}H^c{}_c) + \cdots, \tag{97}$$

$${}^*g_{ta} \sim -\epsilon^3 V_a + \cdots, \tag{98}$$

$${}^*g_{ab} \sim g_{ab} + \tfrac{1}{2}\epsilon^2\psi g_{ab} + \epsilon^4(\tfrac{1}{2}b g_{ab} + H_{ab} - \tfrac{1}{2}H^c{}_c g_{ab}) + \cdots, \tag{99}$$

where here and throughout this paper Latin indices range over r, θ, and φ while Greek indices range over r, θ, φ, and t, and ψ and V_a, etc., are functions of r, θ, φ, and t. See Fig. 3 for an interpretation of the terms in this expansion.

Our task is to pick the functions ψ, V_a, and H_{ab} in such a manner that the resulting stress–energy tensor describes an interesting and realistic physical situation. It is trivial but tedious to compute the Einstein tensor components corresponding to the metric given above. The results of such a computation are

$${}^*G_{tt} \sim -\tfrac{1}{2}\epsilon^2\nabla^2\psi + \cdots, \tag{100}$$

$${}^*G_{ta} \sim \tfrac{1}{2}\epsilon^3[\nabla^2 V_a - (\boldsymbol{\nabla}\cdot\mathbf{V} + \partial_t\psi)_{,a}] + \cdots, \tag{101}$$

$${}^*G_{ab} \sim \epsilon^4\{-\tfrac{1}{2}\nabla^2 H_{ab} + \tfrac{1}{8}\psi_{,a}\psi_{,b} + \tfrac{1}{4}\psi\psi_{,ab} - g_{ab}(\tfrac{3}{16}\boldsymbol{\nabla}\psi\cdot\boldsymbol{\nabla}\psi + \tfrac{1}{4}\psi\nabla^2\psi) + \tfrac{1}{2}(H^c_{a;c} + V_{a,t})_{;b} + \tfrac{1}{2}(H_b{}^c{}_{;c} + V_{b,t})_{;a} - \tfrac{1}{2}g_{ab}[\boldsymbol{\nabla}\cdot(\boldsymbol{\nabla}\cdot\mathbf{H} + \partial_t\mathbf{V}) + \partial_t(\boldsymbol{\nabla}\cdot\mathbf{V} + \partial_t\psi)]\} + \cdots, \tag{102}$$

where I have used the convention of representing 3-vectors in boldface and 3-dyadics in boldface sans serif.[35] The semicolon denotes a covariant derivative in the background space (flat space but curvilinear coordinates). These results differ from the results of the linearized theory only in the appearance of various

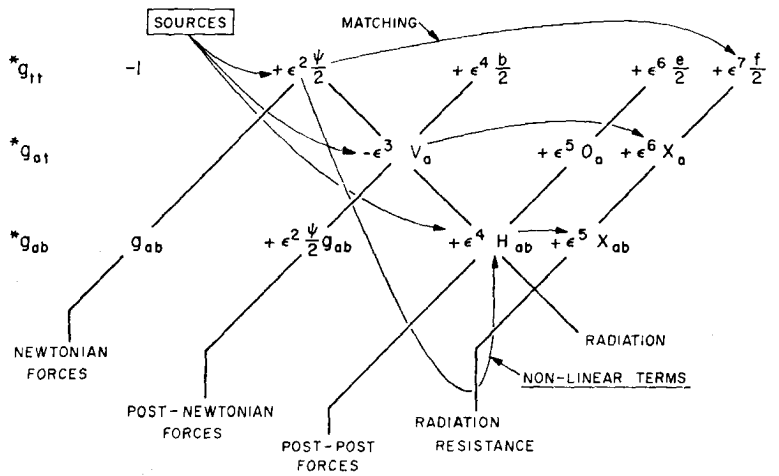

FIG. 3. A schematic diagram of the inner expansion of the metric drawn for $L = 2$. Terms in the expansion that are of the same order in the near zone are placed vertically above each other. The lines of positive slope connect terms which give rise to near-zone forces of the same order. We can see that the vector and tensor parts of the metric are needed to a lower order than the scalar (time–time) part. The lines of negative slope connect terms which contribute to radiation of a given order. The lower multipolarity of the vector and tensor parts increases their size in the wave zone relative to the scalar part. The X_a and X_{ab} are terms removed by our gauge transformation.

squares of the Newtonian potential in the space–space components and in time derivatives being considered small.

The expressions for the components of the Einstein tensor can be simplified considerably if our functions satisfy the auxiliary conditions

$$\nabla \cdot \mathbf{V} + \partial_t \psi = 0, \tag{103}$$

$$\nabla \cdot \mathbf{H} + \partial_t \mathbf{V} = 0, \tag{104}$$

which I refer to as "gauge conditions."

In this case the Einstein field equations

$$G_{\mu\nu} = 8\pi T_{\mu\nu}, \quad (G = 1, \quad c = 1) \tag{105}$$

become

$$\nabla^2 \psi = -16\pi\rho, \tag{106}$$

$$\nabla^2 \mathbf{V} = -16\pi \mathbf{J}, \tag{107}$$

$$\nabla^2 \mathbf{H} = -16\pi[\mathbf{S} + {}^G\mathbf{S}], \tag{108}$$

where

$$\rho \equiv T_{tt}, \quad \text{energy density}, \tag{109}$$

$$J_a \equiv -T_{ta}, \quad \text{momentum flux}, \tag{110}$$

$$S_{ab} \equiv T_{ab}, \quad \text{material stress}, \tag{111}$$

$$^GS_{ab} \equiv (-\tfrac{1}{8}\pi)[\tfrac{1}{8}\psi_{,a}\psi_{,b} + \tfrac{1}{4}\psi\psi_{,ab} - g_{ab}(\tfrac{3}{16}\nabla\psi\cdot\nabla\psi + \tfrac{1}{4}\psi\nabla^2\psi)], \quad \text{gravitational stress}. \tag{112}$$

The gauge conditions (103) and (104), together with the field equations (106)–(108), require the sources to satisfy the conservation laws

$$\nabla \cdot \mathbf{J} + \partial_t \rho = 0, \tag{113}$$

$$\nabla \cdot (\mathbf{S} + {}^G\mathbf{S}) + \partial_t \mathbf{J} = 0, \tag{114}$$

which we recognize as the equations of mass conservation and of momentum conservation ($\mathbf{F} = m\mathbf{a}$). In turn, the above conservation laws imply that if the gauge conditions are satisfied by some initial-value data, then the gauge conditions will be satisfied for all later times. If we do not impose the above conservation laws (force laws), then additional terms appear in $^*G_{ab}$ to compensate.

We can easily compute the "gravitational force"

$$-\nabla \cdot {}^G\mathbf{S} = +\tfrac{1}{4}\rho\nabla\psi. \tag{115}$$

This ψ is related to the Newtonian potential U by

$$\psi = 4U, \tag{116}$$

as can be seen from Eq. (106). Thus we have recovered Newtonian gravity from Einstein's general relativity.

This formalism that we have described has an invariance similar to the gauge invariance of electromagnetism. In gravity it arises from the possibility of describing a physical system in different coordinate systems. Many different metrics $^*g_{\mu\nu}$ represent the same physical system. Expressed in terms of our space–time split, the transformation

$$\psi \to \tilde{\psi} = \psi + \nabla \cdot \boldsymbol{\chi} - \partial_t \chi, \tag{117}$$

$$\mathbf{V} \sim \tilde{\mathbf{V}} = \mathbf{V} + \nabla\chi - \partial_t \boldsymbol{\chi}, \tag{118}$$

$$H_{ab} \to \tilde{H}_{ab} = H_{ab} + \chi_{a;b} + \chi_{b;a} - g_{ab}(\nabla \cdot \boldsymbol{\chi} + \partial_t \chi) \tag{119}$$

leads to a new metric having the same distribution of matter at lowest order, although not necessarily one satisfying the gauge conditions (103) and (104).

B. The Calculation of the Radiation

The electromagnetic radiation in Lorentz gauge was dominated by Y_{LM} in the scalar potential and by $\mathbf{Y}_{L,L-1,M}$ in the vector potential for some value of L. The radiation satisfied the Lorentz gauge condition, and the $\mathbf{Y}_{L,L-1,M}$ multipole moment of the current distribution was therefore determined by the Y_{LM} multipole moment of the mass distribution. In fact, we have

$$a_{LM}(t) = [L(2L+1)]^{-\frac{1}{2}} \frac{dq_{LM}}{dt}, \tag{120}$$

where

$$q_{LM}(t) \equiv \int Y^*_{LM} \rho r^L \, dV, \tag{121}$$

$$a_{LM}(t) \equiv \int \mathbf{Y}^*_{L,L-1,M} \cdot \mathbf{J} r^{L-1} \, dV, \tag{122}$$

which follows directly from the law of conservation of current through integration by parts.

In gravitational radiation a similar situation arises. The equation of continuity, Eq. (113), leads to the same relations (120)–(122) connecting mass and momentum multipoles. In addition, the conservation law for momentum, Eq. (114), determines the stress multipole appearing in the radiation:

$$d_{LM}(t) = [(L-1)(2L-1)]^{-\frac{1}{2}} \frac{da_{LM}}{dt}, \tag{123}$$

where

$$d_{LM}(t) \equiv \int \mathbf{T}^*_{L,L-2,M} \cdot (\mathbf{S} + {}^{G}\mathbf{S}) r^{L-2} \, dV. \tag{124}$$

Thus, in the slow-motion limit the radiation fields have the asymptotic behavior

$$\psi \to Y_{LM} f(r, t), \tag{125}$$

$$\mathbf{V} \to \mathbf{Y}_{L,L-1,M} g(r, t), \tag{126}$$

$$\mathbf{H} \to \mathbf{T}_{L,L-2,M} h(r, t), \tag{127}$$

and the dominant part of the radiation is determined completely by the mass multipole moment $q_{LM}(t)$. There is no change in the radiation caused by a prescribed mass multipole distribution as one goes from the very-weak-field limit, where the dominant stresses are mechanical stresses $\mathbf{S}$, to the post-Newtonian limit, where the dominant stresses contain some gravitational stresses ${}^{G}\mathbf{S}$, since it is the sum of these stresses which appears in both the conservation law (114) and the source equation for the radiation (108). One must, of course, include gravitational forces in determining the motion of post-Newtonian systems, whereas one can neglect it for very-weak-field systems. Since the damping comes directly from the radiation [cf. Eqs. (76) and (77) in the electromagnetic example], it too will be determined if one knows only $q_{LM}(t)$.

C. The Calculation of the Damping

Our formalism of slow-motion gravity is nearly identical to classical electromagnetism. Aside from the additional dyadic field, one can nearly copy the electromagnetic calculations line for line. That is why we paid so much attention to the electromagnetic problem. Here we can now concentrate on the results and their interpretation.

In electromagnetism the near-zone source equation is

$$\nabla^2 \varphi = -4\pi\rho, \tag{128}$$

while in the gravitational case, Eq. (106), it is

$$\nabla^2(\tfrac{1}{4}\psi) = -4\pi\rho. \tag{129}$$

The dominant force in the slow-motion (Newtonian) limit comes from the scalar potential. The electromagnetic force law is

$$\mathbf{F} = -q\nabla\varphi, \tag{130}$$

while the gravitational force law (115) is

$$\mathbf{F} = m\nabla(\tfrac{1}{4}\psi). \tag{131}$$

From this we see that both theories have inverse square static forces, although in gravity likes attract likes because of the sign difference.

In both theories we calculate the dominant part of the near-zone field from the source equations (128) or (129); we then match this to the waves of the radiation zone; and we finally match those waves back into the near zone to get the damping terms associated with each multipole. We then use gauge transformations to put the damping terms into new forms. In these new forms the damping force is determined to lowest order by the leading time-odd part of the transformed scalar potential. The vector and tensor potentials contribute only in higher order. The gauge transformations, along with tensor spherical harmonics and multipole solutions, are discussed in more detail in the Appendix. The result for electromagnetism (electric-parity modes) is

$$\tilde{\varphi} = -[(L+1)/L]\varphi, \tag{132}$$

while for gravity it is

$$\tilde{\psi} = [(L+1)(L+2)/(L-1)L]\psi, \tag{133}$$

where L is the multipole index, $\tilde{\varphi}$ and $\tilde{\psi}$ are the gauge-transformed scalar potentials, and φ and ψ are the untransformed ones.

The sign difference between these equations cancels the sign difference in the force law, and so it is clear that *any gravitational multipole has a resistive field such as we found for electromagnetism* [see Eqs. (82)–(84)], *but which is* $(L+2)/(L-1)$ *times as strong as*

the corresponding electromagnetic resistive field. Since electromagnetism always damps, so must gravity.

From the above outline we can see that the gravitational calculation reduces to a calculation of the factor $(L+1)(L+2)/(L-1)L$ appearing in the gauge transformation (133). This calculation is discussed in more detail in the Appendix.

D. Results

The slow-motion formalism presented in this paper differs from the standard slow-motion work by introducing additional fields, called resistive fields. These fields have potentials which satisfy homogeneous equations and which diverge for large distances. Their introduction is not arbitrary, but is required if the solutions in the near zone are to match properly with purely outgoing radiation in the wave zone. Their divergence for large radius does not reflect a physical divergence, but only indicates the nonuniformity of our inner expansion for large radius. These resistive fields can be calculated directly from the Newtonian motions, as can the radiation.

These resistive fields act on all of the matter in the induction zone. The resistive force $\mathbf{F}_R$ acting per unit volume on the mass density ρ as a result of the gravitational radiation emitted by the changing mass multipole moment $q_{LM}(t)$ is given by

$$\mathbf{F}_R = (-)^{L+1}4\pi\left(\frac{(L+1)(L+2)}{L(L-1)}\right)\frac{[(2L+1)L]^{\frac{1}{2}}}{[(2L+1)!!]^2}\rho \times \sum_M \frac{d^{(2L+1)}q_{LM}(t)}{dt^{(2L+1)}} r^{L-1}\mathbf{Y}_{L,L-1,M}, \quad (134)$$

where the multipole moment is defined by

$$q_{LM}(t) \equiv \int Y^*_{LM} r^L \rho(r, \Omega, t)\, dV. \quad (135)$$

This is calculated by taking the solution to Eq. (106) corresponding to the LM multipole moment, multiplying it by the gauge transformation factor [see Eq. (A36)] and taking the leading time-odd part of the radial dependence [see Eq. (A22)]. This is just $(L+2)/(L-1)$ times the corresponding electromagnetic expression. The $L=2$ case is the only one with resistive forces large enough to be of physical interest. In that case[36]

$$\mathbf{F}_R = -\frac{8\pi\sqrt{10}}{75}\rho r \sum_M \mathbf{Y}_{21M}\frac{d^5 q_{2M}}{dt^5}. \quad (136)$$

The direction of this resistive force is such as to lead to an extraction of energy from the source. The work done by the force is given by

$$\frac{dE}{dt} = \int \mathbf{v}\cdot\mathbf{F}_R\, dV \propto \int \mathbf{J}\cdot\mathbf{Y}_{L,L-1,M} r^{L-1}\, dV, \quad (137)$$

which is proportional to the $(L, L-1, M)$ vector multipole moment of the mass current. This is determined by the mass multipole moment [cf. Eq. (121)]. Using Eqs. (135), (137), and (121) and adding some perfect derivatives to our expression, we find that the power extracted by the resistive field is

$$\frac{dE}{dt} = -\frac{4\pi(L+1)(L+2)}{L(L-1)[(2L+1)!!]^2} \times \sum_M \left(\frac{d^{(L+1)}q_{LM}}{dt^{(L+1)}}\right)^2 + \frac{dX}{dt}, \quad (138)$$

where X is some function of q_{LM} and its derivatives. For periodic (or bounded) motions the time derivative term averages to zero, and energy is extracted from the source. Remember that here L is the first nonzero, time-dependent multipole moment of the mass distribution.

For the special case of harmonic motions we define an amplitude $\hat{q}$ by the equation

$$q_{LM}(t) = \hat{q}_{LM}\cos\omega t. \quad (139)$$

In this case the mean energy lost per radian is

$$\frac{1}{\omega}\left\langle\frac{dE}{dt}\right\rangle = -\frac{2\pi(L+1)(L+2)\omega^{2L+1}}{L(L-1)[(2L+1)!!]^2}\sum_M(\hat{q}_{LM})^2, \quad (140)$$

which for $L=2$ becomes

$$\left\langle\frac{dE}{dt}\right\rangle = -\left(\frac{4\pi\omega^6}{75}\right)\sum_M(\hat{q}_{2M})^2. \quad (141)$$

A useful form of this expression can be found by introducing a typical length l and a typical mass M such that

$$\sum_M(\hat{q}_{2M})^2 = \beta^2 M^2 l^4, \quad (142)$$

where β is a numerical factor of order unity. In this case we can write the energy lost per radian as

$$\frac{\langle dE/dt\rangle}{\omega} = -\frac{4\pi\beta^2}{75}\frac{GM^2}{l}\left(\frac{2\pi l}{\lambda}\right)^5. \quad (143)$$

The second term in this expression is an energy of the size of the gravitational binding energy. Thus we see that the system radiates a small fraction of ($\sim\epsilon^5$) its gravitational binding energy per radian. All of the energy-loss formulas presented here agree with those derived from the Landau–Lifshitz pseudotensor using the elementary linear theory.[37]

For completeness we can also consider the magnetic-parity multipoles. In cases of high symmetry, but not in general, these can compete at lowest order with the electric-parity modes. The same type of reduction can be carried through to show that the gravitational damping of these modes is just $L/(L+1)$ times as strong as the electromagnetic damping of the corresponding mode.

If we define a mass-current multipole moment $d_{LM}(t)$ by

$$d_{LM} \equiv \int \mathbf{Y}^*_{LLM} \cdot \mathbf{J} r^L \, dV, \tag{144}$$

then the resistive fields turn out to be[31]

$$\mathbf{F}_R = (-)^{L+1} 4\pi L(L+1)^{-1}[(2L+1)!!]^{-2} \times \sum_M d^{(2L+2)}_{LM}(t) r^L \mathbf{Y}_{LLM}, \tag{145}$$

and the power extracted per radian for harmonic motions is

$$\frac{\langle dE/dt \rangle}{\omega} = -8\pi L(L+1)^{-1} \times [(2L+1)!!]^{-2} \omega^{2L+1} \sum_M (\hat{d}_{LM})^2. \tag{146}$$

E. Interpretation

The procedures outlined above have been carefully chosen to avoid several sources of confusion. These pitfalls will be mentioned here so that those interested in extending the results will not have to waste time on them.

One perhaps obvious point is that the formulas are only valid for the first multipole moment of a given parity which has a nonzero time derivative. Otherwise, the small corrections to the lower multipoles would lead to errors as large as the effects of the higher multipoles. Also, there may be radiation from post-Newtonian corrections to the lower multipoles (such as kinetic energy corrections to masses) which should be included. This is often the case with magnetic-parity modes.

Another point is that one cannot always compute the resistive force from the force law

$$\mathbf{F}_R = \tfrac{1}{4} m \nabla \psi_R. \tag{147}$$

Before the gauge transformation was used to simplify the problem, the lower L dependence of the vector and tensor potentials allowed them to compete with the scalar potential, and one had to use a more complete force law (derivable by writing the equations for a geodesic of ${}^*g_{\mu\nu}$ in terms of $g_{\mu\nu}$ and $h_{\mu\nu}$).

An important point concerns the physical interpretation of these results. The integration of the equations of motion resulting from the "pseudoforce" law allows one to write down the development of the system referred to some coordinate system. The big difference between gravity and electromagnetism now appears. The electromagnetic field produces effects only through its force law. On the other hand, not only does the gravitational field affect the coordinate motion of the system, but the potentials themselves determine the clock rates and the behavior of rigid bodies. Until one knows the **H** field, one cannot convert coordinate differences into proper length without making errors that are $O(\kappa)$.[38] For the most part we will not need amplitudes any more accurately than $O(1)$. However, there are situations where such an error can make a big effect, and this should be kept in mind.

F. Conclusion

This paper has developed the methods needed to include the irreversible escape of radiation in a slow-motion expansion. The energy lost in the escaping radiation is calculated by finding the work done on the radiating system. The calculation was done for field strengths appropriate to gravitationally bound systems, rather than the infinitely weak fields of the linearized theory, and the formulas of the linearized theory were found to apply to both cases.

ACKNOWLEDGMENTS

I would like to thank K. Thorne, P. Lagerstrom, and F. Estabrook. They helped me greatly with this research and also with its preparation for publication.

APPENDIX: TENSOR SPHERICAL HARMONICS AND MULTIPOLE RADIATION

This appendix contains material useful for solving the tensor wave equation in spherical coordinates.

When studying radiation from confined systems, we can introduce a great deal of simplification by separating the sources and the solutions into "multipoles," having simple behavior under rotations. The ordinary wave equation does not mix the multipoles, and for slow motions only a few multipoles will dominate the radiation field.

The angular functions that will be defined here have the following features. They behave simply under rotations, forming irreducible representations of the rotation group. They have simple formulas for their gradients, curls, and divergences. Finally, they can be computed with relative ease.

The material in this appendix is presented, not because it is original, but because it is relatively inaccessible. Tensor spherical harmonics were developed from the rotation group by Mathews.[39]

Throughout this paper I have followed the conventions of Edmonds,[32] and his book contains all that is needed to supplement the material presented here.

A. General Features

We will define special tensors $\mathbf{T}_{JLM}$, having properties similar to the vector spherical harmonics $\mathbf{Y}_{JLM}$.[40] These $\mathbf{T}_{JLM}$ will be called tensor spherical harmonics and are simultaneous eigenfunctions of S^2, where S_x, S_y, and S_z are the generators of rotations for a second-rank tensor, of L^2, where L_x, L_y, and L_z are the generators of the rotations of a function of spatial variables, and of J^2 and J_z, where

$$\mathbf{J} = \mathbf{L} + \mathbf{S}. \tag{A1}$$

Thus

$$L^2\mathbf{T}_{JLM} = L(L+1)\mathbf{T}_{JLM}, \tag{A2}$$

$$J^2\mathbf{T}_{JLM} = J(J+1)\mathbf{T}_{JLM}, \tag{A3}$$

$$J_z\mathbf{T}_{JLM} = M\mathbf{T}_{JLM}. \tag{A4}$$

These tensor spherical harmonics are combinations of constant basis tensors and scalar spherical harmonics. Thus the Laplacian of the tensor field $\varphi(r,t)\mathbf{T}_{JLM}(\Omega)$ depends only upon the L value (spatial dependence):

$$\nabla^2(\varphi\mathbf{T}_{JLM}) = [r^{-1}\partial_r^2(r\varphi) - L(L+1)r^{-2}\varphi]\mathbf{T}_{JLM}. \tag{A5}$$

When forming tensor multipole solutions, we can use the same radial functions as are used with scalar multipoles and vector multipoles.

B. The Radial Equation

Let us define an operator W_L by

$$W_L \equiv r^{-1}\partial_r^2(r\varphi) - L(L+1)r^{-2}\varphi - \partial_t^2\varphi. \tag{A6}$$

We are interested in homogeneous solutions satisfying

$$W_L\varphi = 0. \tag{A7}$$

One can verify[41] that if φ_L is a solution, that is

$$W_L\varphi_L = 0, \tag{A8}$$

then

$$W_{L+1}(\partial_r - L/r)\varphi_L = 0 \tag{A9}$$

and

$$W_{L-1}[\partial_r + (L+1)/r]\varphi_L = 0. \tag{A10}$$

These relations give us raising and lowering operators for the solutions φ_L. We abbreviate these

$$D_L^+ \equiv \partial_r - L/r, \tag{A11}$$

$$D_L^- \equiv \partial_r + (L+1)/r. \tag{A12}$$

The solution of Eq. (A7) for $L = 0$ is

$$\varphi_0 = F(t \mp r)/r, \tag{A13}$$

where F can be any function of one variable.

Solutions of higher L values may be obtained by using D^+ repeatedly. We have, for example,

$$\varphi_1 = \frac{F'(t \mp r)}{r} \pm \frac{F(t \mp r)}{r^2}, \tag{A14}$$

$$\varphi_2 = \frac{F''(t \mp r)}{r} \pm \frac{3F'(t \mp r)}{r^2} + \frac{3F(t \mp r)}{r^3} \tag{A15}$$

(normalizing the $1/r$ term). We can conveniently abbreviate these radial solutions by writing just the leading term and the L value. Thus we write the $L = 1$ solution

$$\varphi_1 \leftrightarrow \{F'(t \mp r)\}_1, \tag{A16}$$

and our raising and lowering operators applied to these radial functions yield new radial functions according to

$$D_L^+\{F\}_L = \mp\{F'\}_{L+1}, \tag{A17}$$

$$D_L^-\{F\}_L = \mp\{F'\}_{L-1}. \tag{A18}$$

These raising and lowering operators will appear throughout the formulas for the gradient, divergence, etc., of our spherical harmonics.

These functionals $\{F\}$ have been normalized by the size of their radiation (that is, the size of their $1/r$ terms). We will also be interested in their behavior for small r. This can be found from a Taylor's series expansion of

$$\{F^{(p+L)}(t \mp r)\}_L = (\mp)^L\left(\prod_{q=0}^{L-1} D_q^+\right)\{F^{(p)}(t \mp r)\}_0, \tag{A19}$$

which gives us

$$\{F^{(Q)}(t \mp r)\}_L = \sum_{p=0}^{\infty}[(p-1)(p-3)\cdots(p-2L+1)] \times r^{(p-L-1)}\frac{F^{(p+Q-L)}(t \mp r)}{p!}. \tag{A20}$$

The leading term in the small r limit is the $p = 0$ term

$$\{F^{(Q)}\}_L \to (\pm)^L(2L-1)!!\, r^{-L-1}F^{(Q-L)}(t), \tag{A21}$$

and the leading time-odd term in the small r limit is the $p = 2L+1$ term

$$[\{F^{(Q)}\}_L]_R \to (\mp)^{L+1}r^L\frac{F^{(L+Q+1)}(t)}{(2L+1)!!}. \tag{A22}$$

This term is time-odd relative to the induction field, $p = 0$ term, and determines the time-odd effects in the small r limit.

C. Tensor Spherical Harmonics

The general tensor is not an irreducible representation of spin 2 at a point. The trace behaves like a scalar under rotations and the antisymmetric part like a vector. The spin-2 part of a tensor is its traceless, symmetric part.

We can easily generate a suitable set of basis tensors by combining the basis vectors used for the vector spherical harmonics,

$$\mathbf{e}_0 = \mathbf{e}_z, \quad \mathbf{e}_{\pm 1} = (\mp \mathbf{e}_x - i\mathbf{e}_y)/\sqrt{2}. \tag{A23}$$

We define five basis tensors $\mathbf{t}_p, p = -2, -1, \cdots, +2$, by

$$\mathbf{t}_p \equiv \sum_{ss'} (1s1s' \mid 112p)\mathbf{e}_s\mathbf{e}_{s'}, \tag{A24}$$

where $(j_1 m_1 j_2 m_2 \mid j_1 j_2 jm)$ is a Clebsch–Gordan coefficient.

Using these basis tensors, we can construct tensor fields that are suitable eigenfunctions, taking

$$\mathbf{T}_{JLM} \equiv \sum_{SS'} (LS2S' \mid L2JM) Y_{LS}\mathbf{t}_{S'}. \tag{A25}$$

Perhaps the most useful formulas for these tensor fields relate the divergence of a tensor spherical harmonic to vector spherical harmonics. One can write this in the form

$$\begin{aligned}\nabla \cdot (\varphi \mathbf{T}_{JLM}) &= (-)^{J+L+1}5^{\frac{1}{2}} \\ &\times \left(-(L+1)^{\frac{1}{2}} \begin{Bmatrix} L+1 & 1 & L \\ 2 & J & 1 \end{Bmatrix} \mathbf{Y}_{J,L+1,M} D_L^+ \varphi \right. \\ &\left. + L^{\frac{1}{2}} \begin{Bmatrix} L-1 & 1 & L \\ 2 & J & 1 \end{Bmatrix} \mathbf{Y}_{J,L-1,M} D_L^- \varphi \right),\end{aligned} \tag{A26}$$

where the symbol

$$\begin{Bmatrix} j_1 & j_2 & j_{12} \\ j_3 & j & j_{23} \end{Bmatrix}$$

is a 6-j symbol. These can be either looked up[42] or calculated from the formulas given in Edmonds.

In discussing the gauge transformations useful in weak-field gravity, we will need the symmetrized gradients of our vector spherical harmonics expressed as tensor spherical harmonics. We define these by

$$(\nabla \mathbf{V} + \mathbf{V}\nabla)_{ab} \equiv \partial_a V_b + \partial_b V_a, \quad a, b = x, y, z. \tag{A27}$$

A straightforward computation ("addition of 3 angular momenta") gives us

$$\begin{aligned}(\nabla \varphi \mathbf{Y}_{JLM} + \varphi \mathbf{Y}_{JLM}\nabla) - \tfrac{2}{3}\mathbf{I}(\nabla \cdot \varphi \mathbf{Y}_{JLM}) &= (-)^{J+L+1}2(5)^{\frac{1}{2}} \\ &\times \left(-(L+1)^{\frac{1}{2}} D_L^+ \varphi \begin{Bmatrix} L+1 & 1 & L \\ 1 & J & 2 \end{Bmatrix} \mathbf{T}_{J,L+1,M} \right. \\ &\left. + L^{\frac{1}{2}} D_L^- \varphi \begin{Bmatrix} L-1 & 1 & L \\ 1 & J & 2 \end{Bmatrix} \mathbf{T}_{J,L-1,M} \right),\end{aligned} \tag{A28}$$

where

$$(\mathbf{I})_{ab} \equiv g_{ab}. \tag{A29}$$

D. Multipole Fields

By combining the tensor spherical harmonics defined in Eq. (A25) with the solutions of the radial equation constructed in subsection B of this appendix, we can easily write down multipole solutions of the weak-field equations (106)–(108).

Solutions having "electric parity" may be written as

$$\mathbf{H} = [(2L+1)(2L-1)/L(L-1)]^{\frac{1}{2}} \times \{F^{(L)}(t \mp r)\}_{L-2}\mathbf{T}_{L,L-2,M}, \tag{A30}$$

$$\mathbf{V} = \pm[(2L+1)/L]^{\frac{1}{2}}\{F^{(L)}(t \mp r)\}_{L-1}\mathbf{Y}_{L,L-1,M}, \tag{A31}$$

$$\psi = \{F^{(L)}(t \mp r)\}_L Y_{LM} \tag{A32}$$

or as

$$\tilde{H}_{ab} = [(2L+1)(2L+3)/(L+1)(L+2)]^{\frac{1}{2}} \times \{G^{(L)}(t \mp r)\}_{L+2}(\mathbf{T}_{L,L+2,M})_{ab}, \tag{A33}$$

$$\tilde{\mathbf{V}} = \mp[(2L+1)/(L+1)]^{\frac{1}{2}} \times \{G^{(L)}(t \mp r)\}_{L+1}\mathbf{Y}_{L,L+1,M}, \tag{A34}$$

$$\tilde{\psi} = \{G^{(L)}(t \mp r)\}_L Y_{LM}. \tag{A35}$$

These will be different representations of the same solution and related by a gauge transformation [Eqs. (117)–(119)], provided that

$$\tilde{\psi} = [(L+1)(L+2)/L(L-1)]\psi. \tag{A36}$$

This is calculated by writing the gauge functions $\boldsymbol{\chi}$ and χ as multipoles of appropriate J value and parity, taking the radial functions to be the appropriate radial solutions (A20), using the formulas for spherical harmonics, especially Eq. (A28), to differentiate these terms, and finally picking the coefficients of the terms of $\boldsymbol{\chi}$ and χ so that the transformed fields have the form of Eqs. (A33)–(A35).

"Magnetic-parity" solutions (i.e., those which couple to mass currents) may be written as

$$\mathbf{H} = \pm[2(2L+1)/(L-1)]^{\frac{1}{2}} \times \{F^{(L)}(t \mp r)\}_{L-1}\mathbf{T}_{L,L-1,M}, \quad \text{(A37)}$$

$$\mathbf{V} = \{F^{(L)}(t \mp r)\}_L Y_{LLM}, \quad \text{(A38)}$$

$$\psi = 0 \quad \text{(A39)}$$

or as

$$\tilde{H}_{ab} = \mp[2(2L+1)/(L+2)]^{\frac{1}{2}} \times \{G^{(L)}(t \mp r)\}_{L+1}(\mathbf{T}_{L,L+1,M})_{ab}, \quad \text{(A40)}$$

$$\tilde{\mathbf{V}} = \{G^{(L)}(t \mp r)\}_L \mathbf{Y}_{LLM}, \quad \text{(A41)}$$

$$\tilde{\psi} = 0, \quad \text{(A42)}$$

and these will be representations of the same solution provided that

$$\tilde{\mathbf{V}} = -[L/(L+1)]\mathbf{V}. \quad \text{(A43)}$$

* Supported in part by the National Science Foundation and the Office of Naval Research.

† Present address: Lick Observatory, University of California, Santa Cruz, Calif.

[1] For more detail, see J. Cole, *Perturbation Methods in Applied Mathematics* (Ginn-Blaisdell, New York, 1968).

[2] J. N. Goldberg, "The Equations of Motion," in *Gravitation: An Introduction to Current Research*, edited by L. Witten (Wiley, New York, 1962), pp. 102–29.

[3] F. A. E. Pirani, "Gravitational Radiation," in Ref. 2, pp. 199–226.

[4] W. B. Bonnor, Brit. J. Appl. Phys. **14**, 555 (1963).

[5] A. Trautman, Usp. Fiz. Nauk **89**, 3 (1966) [Sov. Phys. Usp. **9**, 319 (1966)].

[6] L. Infeld and J. Plebanski, *Motion and Relativity* (Pergamon, New York, 1960).

[7] A. Einstein and L. Infeld, Can. J. Math. **1**, 209 (1949).

[8] V. Fock, *The Theory of Space, Time and Gravitation* (Macmillan, New York, 1964), 2nd ed., pp. 398, 399.

[9] A. Einstein, L. Infeld, and B. Hoffman, Ann. Math. **39**, 65 (1938).

[10] S. Chandrasekhar, Astrophys. J. **142**, 1488, 1513 (1965); **158**, 45 (1969).

[11] S. Chandrasekhar and Y. Nutku, Astrophys. J. **158**, 55 (1969).

[12] S. Chandrasekhar and P. Esposito, Astrophys. J. **160**, 153 (1970).

[13] A. Trautman, Bull. Acad. Polon. Sci. Ser. Sci. Math. Astron. Phys. **6**, 627 (1958).

[14] A. Peres, Nuovo Cimento **15**, 351 (1960).

[15] A. Peres and N. Rosen, Ann. Phys. **10**, 94 (1960).

[16] M. Carmeli, Nuovo Cimento **37**, 842 (1965).

[17] L. Infeld and R. Michalska-Trautman, Ann. Phys. **55**, 561, 576 (1969).

[18] P. Havas, Phys. Rev. **108**, 1351 (1957).

[19] P. Havas and J. N. Goldberg, Phys. Rev. **128**, 398 (1962).

[20] S. F. Smith and P. Havas, Phys. Rev. **138**, B495 (1965).

[21] W. B. Bonnor, Phil. Trans. Roy. Soc. London **A251**, 233 (1959).

[22] W. B. Bonnor and M. A. Rotenberg, Proc. Roy. Soc. (London) **A289**, 247 (1965).

[23] Reference 1.

[24] The singularity at $x = 0$ present in this example is not an essential feature of a "singular" problem. The argument could have been pursued with a function such as $f(x, \epsilon) = \cos x - \exp(x/\epsilon)$ at the cost of complicating Fig. 1.

[25] S. Kaplun, *Fluid Mechanics and Singular Perturbations*, edited by P. A. Lagerstrom, L. N. Howard, and C. Liu (Academic, New York, 1967).

[26] Actually, we have here a very simple analog of the "Stokes' paradox." In 1851 Stokes proposed a simplified set of equations for viscous flow past a finite object, valid for the case of very low Reynold's number. For 2-dimensional flow, all nontrivial solutions of Stokes' equations diverge logarithmically at infinity. This "paradox" was resolved by Kaplun, who pointed out in 1954 that Stokes' equations are only inner equations and their solutions are not valid far from the body. See Ref. 25.

[27] T. L. Saaty, *Modern Nonlinear Equations* (McGraw-Hill, New York, 1967), Chap. 5.

[28] The spurious solutions can be eliminated easily if the equations are solved by the method of two-time scales. Before a numerical solution is attempted, one should eliminate the higher derivative terms by differentiating the approximate equation and using that to eliminate the third derivative, etc.

[29] Runaway solutions also appear in some mechanical systems, for example, in the case of the spherically symmetric pulsations of a sphere which as a result radiates monopole sound waves. A more complete exposition of this material can be found in Phys. Rev. A **2**,1501 (1970).

[30] Conditions under which singular problems arise are discussed in M. Van Dyke, *Perturbation Methods in Fluid Mechanics* (Academic, New York, 1964), p. 80.

[31] W. L. Burke, "The Coupling of Gravitational Radiation to Non-Relativistic Sources," California Institute of Technology, 1969, available from University Microfilms, Ann Arbor, Michigan.

[32] For these, see A. R. Edmonds, *Angular Momentum in Quantum Mechanics* (Princeton U.P., Princeton, N.J., 1960).

[33] This paper uses the sign conventions recommended by C. W. Misner, K. S. Thorne, and J. A. Wheeler ("Open Letter to Relativity Theorists," August 19, 1968), with signature +2.

[34] This κ is quite similar to the coupling parameter κ which appeared in the string problem. The reader might enjoy working out the slow-motion approximation for the spring-string system, relaxing the requirement that $\kappa \ll \epsilon$.

[35] We have, for example, $(\mathbf{V})_a = V^{at} = -V_{at}$, and will always take the contravariant components as our 3-vectors.

[36] Since ρ is real we have $q^*_{LM} = (-)^M q_{L,-M}$, thus the F defined in Eq. (136) is real as is dE/dt.

[37] See, for example, L. D. Landau and E. M. Lifshitz, *The Classical Theory of Fields* (Addison-Wesley, Reading, Mass., 1962).

[38] Warning! Remember that we have calculated only the leading time-odd term in H. There are much larger time-even, post-Newtonian, and post-post-Newtonian corrections that have been ignored.

[39] J. Mathews, J. Soc. Ind. Appl. Math. **10**, 768 (1962).

[40] See Ref. 32.

[41] Use $[W_L, \partial_r] = (2/r^2)\partial_r - 2L(L+1)/r^3$ and $[W_L, 1/r] = -(2/r^2)\partial_r$.

[42] M. Rotenberg *et al.*, *The 3-j and 6-j Symbols* (MIT, Cambridge, 1959).

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 12, NUMBER 3 MARCH 1971

Solution of the One-Dimensional N-Body Problems with Quadratic and/or Inversely Quadratic Pair Potentials

F. CALOGERO*
Institute of Theoretical and Experimental Physics, Moscow, USSR

(Received 27 May 1970)

The quantum-mechanical problems of N 1-dimensional equal particles of mass m interacting pairwise via quadratic ("harmonical") and/or inversely quadratic ("centrifugal") potentials is solved. In the first case, characterized by the pair potential $\frac{1}{4}m\omega^2(x_i - x_j)^2 + g(x_i - x_j)^{-2}$, $g > -\hbar^2/(4m)$, the complete energy spectrum (in the center-of-mass frame) is given by the formula

$$E = \hbar\omega(\tfrac{1}{2}N)^{\frac{1}{2}}\left[\tfrac{1}{2}(N-1) + \tfrac{1}{2}N(N-1)(a+\tfrac{1}{2}) + \sum_{l=2}^{N} l n_l\right],$$

with $a = \frac{1}{2}(1 + 4mg\hbar^{-2})^{\frac{1}{2}}$. The $N-1$ quantum numbers n_l are nonnegative integers; each set $\{n_l; l = 2, 3, \cdots, N\}$ characterizes uniquely one eigenstate. This energy spectrum can also be written in the form $E_s = \hbar\omega(\frac{1}{2}N)^{\frac{1}{2}}[\frac{1}{2}(N-1) + \frac{1}{2}N(N-1)(a+\frac{1}{2}) + s]$, $s = 0, 2, 3, 4, \cdots$, the multiplicity of the sth level being then given by the number of different sets of $N-1$ nonnegative integers n_l that are consistent with the condition $s = \sum_{l=2}^{N} l n_l$. These equations are valid independently of the statistics that the particles satisfy, if $g \neq 0$; for $g = 0$, the equations remain valid with $a = \frac{1}{2}$ for Fermi statistics, $a = -\frac{1}{2}$ for Bose statistics. The eigenfunctions corresponding to these energy levels are not obtained explicitly, but they are rather fully characterized. A more general model is similarly solved, in which the N particles are divided in families, with the same quadratic interaction acting between all pairs, but with the inversely quadratic interaction acting only between particles belonging to the same family, with a strength that may be different for different families. The second model, characterized by the pair potential $g(x_i - x_j)^{-2}$, $g > -\hbar^2/(4m)$, contains only scattering states. It is proved that an initial scattering configuration, characterized (in the phase space sector defined by the inequalities $x_i \geq x_{i+1}$, $i = 1, 2, \cdots, N-1$, to which attention may be restricted without loss of generality) by (initial) momenta p_i, $i = 1, 2, \cdots, N$, goes over into a final configuration characterized uniquely by the (final) momenta p_i', with $p_i' = p_{N+1-i}$. This remarkably simple outcome is a peculiarity of the case with equal particles (i.e., equal masses and equal strengths of all pair potentials).

1. INTRODUCTION

The motivation for a physicist to study 1-dimensional problems is best illustrated by the story of the man who, returning home late at night after an alcoholic evening, was scanning the ground for his key under a lamppost; he knew, to be sure, that he had dropped it somewhere else, but only under the lamppost was there enough light to conduct a proper search. In a more serious vein, a motivation is perceived[1] in the insight that exact solutions, even of oversimplified models, may provide, and in the possibility to assess the reliability of approximation techniques that can be used in more realistic contexts, by first testing them in exactly solvable cases. Moreover for some physical problems a 1-dimensional schematization may indeed be appropriate. And a final argument emphasizes the formal elegance that exactly solvable models often reveal.

There do not exist many examples of N-body problems with pair forces that can be solved exactly, even in one dimension. The case with only quadratic ("harmonical") potentials may be reduced, by a linear reshuffle of the particle variables, to a problem with decoupled oscillators; thus it can easily be solved, both in the classical and quantal cases, and in a space with any number of dimensions. However, just because it can be reduced to a problem with decoupled variables, it is not so interesting as an example of many-body dynamics, though of course it is quite important from other points of view (for instance, to provide a denumerable basic set of eigenstates for the description of the system). The only other known solvable example, first introduced by Berezin, Pochil, and Finkelberg and by McGuire, is that of N equal-mass particles interacting in one dimension via 2-body equal-strength zero-range δ-function potentials.[2] This example is very interesting, especially in the case of attractive forces, when a collection of many-body bound states exist.[3] But the zero-range character of the forces, implying that two particles interact only when their positions coincide, reduces the problem, at least in the classical case, to a sequence of 2-body processes, whose outcome is primarily determined by kinematics alone (energy and momentum conservation imply that in a 1-dimensional 2-body collision between equal-mass particles no new momenta can be produced; the two particles either maintain their initial momenta or exchange them). And indeed, as remarked by McGuire, it is just because this simplification is

maintained in the quantal case (for equal particles) that the problem is also solvable in this case.

In this paper we present a rather complete analysis (in the framework of quantum mechanics) of two new 1-dimensional N-body problems that bear some resemblance to the two models mentioned above. The first problem considers N equal particles interacting via pair potentials that are the sum of a quadratic ("harmonical") plus an inversely quadratic ("centrifugal") term:

$$V(x_i - x_j) = \tfrac{1}{4}m\omega^2(x_i - x_j)^2 + g(x_i - x_j)^{-2}, \qquad g > -\hbar^2/(4m). \quad (1.1)$$

For $N = 3$ this problem was completely solved recently, i.e., all its eigenvalues and eigenfunctions were explicitly exhibited.[4] Moreover, in the N-body case, a subset (including the ground-state) of the wavefunctions and energy levels were found, and the conjecture was put forward that, irrespective of the statistics that the particle obey (Boltzmann, Bose, or Fermi), the complete energy spectrum for this problem differ from the spectrum of the corresponding problem with $g = 0$ (i.e., with only harmonical forces) and with Fermi statistics, only by the (N-dependent) constant[5]

$$\Delta_F E = \tfrac{1}{2}N(N-1)\hbar\omega(\tfrac{1}{2}N)^{\frac{1}{2}}\tfrac{1}{2}[(1 + 4mg\hbar^{-2})^{\frac{1}{2}} - 1]. \quad (1.2)$$

This conjecture is validated in the present paper by a computation of the complete spectrum. Indeed we prove that, *irrespective of the statistics (Boltzmann, Bose, or Fermi) that the particles obey, the energy spectrum (in the center-of-mass frame) of the 1-dimensional N-body problem with the pair potential (1.1) (with $g \neq 0$) coincides (except for a constant shift of all energy levels) with the energy spectrum of the corresponding problem with only harmonical forces ($g = 0$) and identical particles (bosons or fermions). The shift has the value $\Delta_F E$, Eq. (1.2), relative to the fermion case, and*

$$\Delta_B E = \tfrac{1}{2}N(N-1)\hbar\omega(\tfrac{1}{2}N)^{\frac{1}{2}}\tfrac{1}{2}[(1 + 4mg\hbar^{-2})^{\frac{1}{2}} + 1] \quad (1.3)$$

relative to the boson case. The coincidence of these spectra (after the shift has been taken into account) refers not only to the values of the energy levels, but also to their multiplicities. This result also implies the coincidence [except for the energy shift

$$-\Delta_F E + \Delta_B E = \tfrac{1}{2}N(N-1)\hbar\omega(\tfrac{1}{2}N)^{\frac{1}{2}}]$$

of the spectra of the 1-dimensional N-body systems with only harmonical forces and Fermi or Bose statistics.[6] It should be emphasized that, even though the spectrum of the N-body system under consideration coincides, except for a constant shift, with that of the N-body system with only harmonical forces and Bose or Fermi statistics, the presence of the additional "centrifugal" potential excludes the possibility of reducing the problem to one with decoupled variables by a simple redefinition of the particle variables.

The second model that we study might be regarded as a special case of the first, obtained eliminating from it the harmonical potential, i.e., setting $\omega = 0$. But, in fact, the two models differ qualitatively, because, in the first case, the energy spectrum (in the center-of-mass frame) is discrete, and only localized states are allowed (corresponding to classical orbits that are restricted to a finite phase-space region; in fact, these classical orbits are presumably all closed[7]), while in the latter case the spectrum is continuous and only scattering states occur. The solution of the scattering problem for this model in the 3-body case was evinced by Marchioro[8] from the stationary eigenfunctions given explicitly in Ref. 4. He proved that, both in the classical and the quantal cases, *an initial "ingoing" scattering configuration, characterized (in the sector of configuration space defined by the inequalities $x_i \geq x_{i+1}$, $i = 1, 2, \cdots, N-1$, to which attention may be confined without loss of generality; see below) by the momenta p_i, $i = 1, 2, \cdots, N$, goes eventually over into a uniquely determined "outgoing" configuration characterized by (final) momenta p_i' determined by the simple rule*

$$p_i' = p_{N+1-i}, \quad i = 1, 2, \cdots, N. \quad (1.4)$$

This remarkable outcome (proved by Marchioro for $N = 3$) coincides with that that obtains in the case of infinitely repulsive δ-function potentials. However, while in the δ-function case this result is, at least in the classical case, quite trivial, because, as mentioned above, the zero-range character of the forces reduces the scattering process to a sequence of 2-body encounters whose outcome is determined by kinematics alone,[2] in the present case, owing to the long range character of the interaction, the simple relation (1.4) between the asymptotic momenta obtains after a quite complicated, definitely nontrivial, time evolution. In this paper we prove that, as conjectured by Marchioro,[8] also in the N-body case the remarkable rule of Eq. (1.4) obtains. The proof is done directly in the quantal case; clearly the validity of the result also in the classical case is implied (but a detailed proof through the explicit solution of the equations of motion would be a nontrivial task). Arguments are

also given that imply that the simple rule (1.4) is characteristic only of the case with equal particles (i.e., with equal masses and equal coupling constants).

Let us finally emphasize the two features of the models considered in this paper that presumably have more potential for applications, especially for testing approximation techniques: the singular nature at zero range of the inversely quadratic potential, and its long range.

The first model is treated in Sec. 2; in Sec. 3, a generalization of it is discussed, where the N particles are divided into families, and there are equal quadratic potentials acting between all particle pairs but inversely quadratic potentials acting only between pairs belonging to the same family, with coupling constants that can be different for different families. In Sec. 4, we treat the second type of model, characterized by inversely quadratic potentials and only scattering states. Finally, in Sec. 5 we mention some queries that are naturally suggested by the quest for a more complete understanding of these types of models and we discuss the prospects of further generalizations. Some material has been relegated in five Appendices. Although we use some results of previous papers (especially of C) without reporting their proofs, the presentation should be sufficiently selfcontained to be understandable by readers not already familiar with the subject.

2. THE SYSTEM WITH QUADRATIC AND INVERSELY QUADRATIC POTENTIALS

The Hamiltonian of the system under consideration is

$$H = -\frac{\hbar^2}{2m}\sum_{i=1}^{N}\frac{\partial^2}{\partial x_i^2} + \sum_{i=2}^{N}\sum_{j=1}^{i-1}\{\tfrac{1}{4}m\omega^2(x_i - x_j)^2 + g(x_i - x_j)^{-2}\}. \quad (2.1)$$

Throughout this paper we assume that

$$g > -\hbar^2/(4m) \quad (2.2)$$

to prevent the collapse that a more attractive inversely quadratic potential would cause.[4,9] (In the classical case the condition becomes of course $g > 0$.) Our task is to solve the eigenvalue equation

$$H\psi_s = E_s\psi_s, \quad (2.3)$$

where ψ_s is a translation invariant eigenfunction; in particular, we shall find all the eigenvalues E_s together with their multiplicities.

Hereafter we restrict attention to the sector of configuration space corresponding to a definite ordering of the particles, say

$$x_i \geq x_{i+1}, \quad i = 1, 2, \cdots, N-1. \quad (2.4)$$

In fact, the singular nature of the centrifugal interaction, together with the restriction to one space dimension, forbids any particle from overtaking any other particle. This is reflected in the vanishing not only of the wavefunction, but also of its derivative with respect to any particle coordinate, whenever the coordinates of two particles coincide.[10] As a consequence the extension of the wavefunction to the whole configuration space is achieved by the simple prescription

$$\psi(Px) = \eta_P\psi(x), \quad (2.5)$$

where x indicates the set $\{x_i; i = 1, 2, \cdots, N\}$ of Eq. (2.4), P indicates an arbitrary permutation, η_P equals unity if the particles obey Bose statistics and equals the parity of the permutation if the particles obey Fermi statistics. If the particles are distinguishable (Boltzmann statistics), each wavefunction ψ may correspond to $N!$ different states, each one of these being characterized by a wavefunction vanishing identically for all but one of the $N!$ possible orderings of the particles, and for that one being given by Eq. (2.5). However, rather than considering these $N!$ states as degenerate eigenstates of the same system, it is more appropriate to view them as $N!$ altogether different systems, the difference being enforced by the dynamical superselection rule that prevents particles from crossing over each other. Therefore, hereafter, when counting the multiplicity of the eigenvalues, we shall assume that, irrespective of the original nature of the particles (identical or distinguishable), one and only one state correspond to each different eigenfunction of Eq. (2.3). Hence, there shall be no need to distinguish between the different statistics, since such a distinction affects neither the spectrum nor its multiplicities, nor indeed the wavefunctions in the sector (2.4), but only the prescription to continue them elsewhere. This situation is of course consistent with the remark that, since the singular centrifugal force prevents the particles from crossing over one another, their sequential order can in fact be used to identify them even if they were undistinguishable to begin with.

We now assert that the normalizable solutions of Eq. (2.3) can be cast into the form

$$\psi(x) = z^{a+\frac{1}{2}}\varphi(r)P_k(x), \quad (2.6)$$

where the variables z and r and the constant a are

defined as in C, namely

$$z = \prod_{i=2}^{N} \prod_{j=1}^{i-1} (x_i - x_j), \tag{2.7}$$

$$r^2 = \frac{1}{N} \sum_{i=2}^{N} \sum_{j=1}^{i-1} (x_i - x_j)^2, \tag{2.8}$$

$$a = \tfrac{1}{2}(1 + 4mg\hbar^{-2})^{\frac{1}{2}}, \tag{2.9}$$

and $P_k(x)$ is a homogeneous polynomial of degree k in the particle coordinates.[11] These polynomials are also assumed to be translation invariant and to be solutions of the "generalized Laplace equation"

$$\left[\sum_{i=1}^{N} \frac{\partial^2}{\partial x_i^2} + 2(a + \tfrac{1}{2}) \sum_{i=2}^{N} \sum_{j=1}^{i-1} (x_i - x_j)^{-1} \times \left(\frac{\partial}{\partial x_i} - \frac{\partial}{\partial x_j}\right)\right] P_k(x) = 0. \tag{2.10}$$

We indicate hereafter with $g(N, k)$ the number of independent polynomials, homogeneous of degree k and translation invariant, that are solutions of this generalized Laplace equation [a nonpositive value of $g(N, k)$ indicates that no translation-invariant polynomial solution of Eq. (2.10) exists for the corresponding values of N and k]. The properties of these polynomials and the values of the quantities $g(N, k)$ are discussed below and in Appendix C.

To prove our assertion we insert the ansatz (2.6) into Eq. (2.3), and using Eq. (2.10) and the homogeneity of $P_k(x)$ we obtain (see Appendix A for a detailed derivation)

$$-\hbar^2/(2m)\{\varphi'' + [N + 2k - 2 + N(N-1)(a + \tfrac{1}{2})]r^{-1}\varphi'\} + \{\tfrac{1}{4}m\omega^2 N r^2 - E\}\varphi = 0, \tag{2.11}$$

where the primes indicate differentiation. The normalizable solutions of this equation are of course the functions

$$\varphi_{nk}(r) = \exp[-\tfrac{1}{2}(m\omega/\hbar)(\tfrac{1}{2}N)^{\frac{1}{2}} r^2] L_n^b[(m\omega/\hbar)(\tfrac{1}{2}N)^{\frac{1}{2}} r^2], \quad n = 0, 1, 2, \cdots, \tag{2.12}$$

where L_n^b is a Laguerre polynomial[12] and

$$b = k + \tfrac{1}{2}(N - 3) + \tfrac{1}{2}N(N-1)(a + \tfrac{1}{2}). \tag{2.13}$$

The corresponding energy eigenvalues are

$$E_{2n+k} = \hbar\omega(\tfrac{1}{2}N)^{\frac{1}{2}} \times [\tfrac{1}{2}(N-1) + \tfrac{1}{2}N(N-1)(a + \tfrac{1}{2}) + 2n + k], \quad n = 0, 1, 2, \cdots, \quad k = 0, 1, 2, \cdots. \tag{2.14}$$

This formula may be rewritten as follows:

$$E_s = E_s^F + \Delta_F E, \quad s = 0, 1, 2, \cdots, \tag{2.15a}$$

$$= E^B + \Delta_B E, \quad s = 0, 1, 2, \cdots, \tag{2.15b}$$

with

$$E_s^F = \hbar\omega(\tfrac{1}{2}N)^{\frac{1}{2}}[\tfrac{1}{2}(N^2 - 1) + s], \quad s = 0, 1, 2, \cdots, \tag{2.16a}$$

$$E_s^B = \hbar\omega(\tfrac{1}{2}N)^{\frac{1}{2}}[\tfrac{1}{2}(N - 1) + s], \quad s = 0, 1, 2, \cdots, \tag{2.16b}$$

and

$$\Delta_F E = \tfrac{1}{2}N(N-1)\hbar\omega(\tfrac{1}{2}N)^{\frac{1}{2}}(a - \tfrac{1}{2}), \tag{2.17a}$$

$$\Delta_B E = \tfrac{1}{2}N(N-1)\hbar\omega(\tfrac{1}{2}N)^{\frac{1}{2}}(a + \tfrac{1}{2}). \tag{2.17b}$$

As shall be proved below, E_s^F, respectively E_s^B, Eqs. (2.16a) and (2.16b), are the energy levels of the 1-dimensional N-body problem with only oscillator forces ($g = 0$) and with Fermi, respectively Bose, statistics. Moreover, the multiplicity of the energy levels E_s, E_s^F, and E_s^B (for the respective problems) is the same, and it is given by the formula (proved in Appendix C)

$$f(N, s) = \sum_n g(N, s - 2n)\theta(s - 2n)\theta(n). \tag{2.18}$$

Here, and always in the following,

$$\theta(x) = 1, \quad x \geq 0, \qquad = 0, \quad x < 0, \tag{2.19}$$

and all sums run over integral values of the dummy indices. A vanishing value of $f(N, s)$ (as it occurs, for instance, for $s = 1$; see below) indicates that the corresponding energy level is not present. A more detailed discussion of $f(N, s)$ is given in Appendix C; in particular, it is shown that $f(N, s)$ is the number of completely symmetrical polynomials that are homogeneous of degree s and translation invariant, and that it also coincides with the number of different solutions of the equation

$$s = \sum_{l=2}^{N} l n_l, \tag{2.20}$$

where n_l are nonnegative integers. Explicit expressions for $f(N, s)$ for N up to 5 and arbitrary s are also reported in Appendix C, together with a proof of the asymptotic formula

$$\lim_{s \to \infty} [f(N, s) s^{2-N}] = [N!\,(N-2)!]^{-1}. \tag{2.21}$$

While these equations provide explicit information on the multiplicity of the energy level E_s, it is often more useful to use for the energy spectrum the formula

$$E_{\{n_l\}} = \hbar\omega(\tfrac{1}{2}N)^{\frac{1}{2}} \times \left[\tfrac{1}{2}(N-1) + \tfrac{1}{2}N(N-1)(a + \tfrac{1}{2}) + \sum_{l=2}^{N} l n_l\right]. \tag{2.22}$$

Here, each of the $N-1$ integers n_i can take any non-negative value, and to each set $\{n_i\}$ there corresponds one and only one eigenstate. Thus, this form of the formula for the spectrum automatically takes care of the multiplicity problem; this is implied by a comparison with the previous formula, Eq. (2.15), and by the statement reported above [see Eq. (2.20)].

The polynomials $P_k(x)$ are, by definition, translation invariant and homogeneous of degree k, and they satisfy the generalized Laplace equation (2.10).[13] *The last requirement implies that they are completely symmetrical under the exchange of any two coordinates* x_i, x_j; a formal proof of this most important property is given in Appendix B. Thus, they can be written in the form

$$P_k(x) = S \sum_{\{n_i\}} a_{\{n_i\}} \left[\prod_{i=1}^{N} x_i^{n_i} \right] \left[\prod_{i=1}^{N} \theta(n_i - n_{i-1}) \right] \delta_{k,\Sigma}, \tag{2.23}$$

this being the most general form that a completely symmetrical polynomial can take. Here, by definition,

$$\Sigma = \sum_{i=1}^{N} n_i, \tag{2.24}$$

S is the operator that symmetrizes over all exchanges of the coordinates, $\{n_i\}$ indicates a set of the N (non-negative) integers n_i, and $\sum_{\{n_i\}}$ indicates the sum over all such sets. The sum extends, in fact, only over the sets satisfying the conditions

$$n_i \geq n_{i-1} \geq n_0 \equiv 0, \quad i = 1, 2, \cdots, N, \tag{2.25}$$

$$k = \sum_{i=1}^{N} n_i, \tag{2.26}$$

these restrictions being explicitly enforced by the θ functions and by the Kronecker-δ function; by convention, $n_0 \equiv 0$. The constants $a_{\{n_i\}}$ must, of course, be chosen so that the polynomial of Eq. (2.23) be translation invariant and satisfy the generalized Laplace equation (2.10); for small values of N and k one can easily find in this manner all the polynomials $P_k(x)$. A different way to write the most general completely symmetrical polynomial of N variables that is homogeneous of degree k is reported in Appendix C.

There remains to prove that the spectra of Eqs. (2.16a) and respectively (2.16b) [or, equivalently, Eq. (2.22) with $a = \frac{1}{2}$, respectively $a = -\frac{1}{2}$] correspond to the cases with only oscillator forces ($g = 0$) and Fermi, respectively Bose, statistics. This follows from the remark that, for $g = 0$, the ansatz of Eq. (2.6) may be used, with $a = \frac{1}{2}$ (Fermi statistics) or $a = -\frac{1}{2}$ (Bose statistics), throughout all configuration space—i.e., not only in the sector specified by Eq. (2.4).[14] In the Fermi case, one obtains thereby again the generalized Laplace equation (2.10) (with $a = \frac{1}{2}$), and the fact that only symmetrical polynomials are solutions of this equation is consistent with the requirement that the wavefunction ψ of Eq. (2.6) be completely antisymmetrical [note that z is completely antisymmetrical, while of course r, and therefore also any function $\varphi(r)$, is completely symmetrical]. In the Bose case, the ansatz of Eq. (2.6) with $a = -\frac{1}{2}$ implies that the (homogeneous and translation-invariant) polynomials $P_k(x)$ satisfy the usual Laplace equation, namely Eq. (2.10) with $a = -\frac{1}{2}$; though nonsymmetrical polynomial solutions of this equation do now exist, they are excluded by the symmetry requirement itself, and this guarantees that also the multiplicity for the Bose case coincides with that of the general case (the number of completely symmetrical polynomial solutions, homogeneous of degree k and translation invariant, of the Laplace equation, coincides with the number of polynomial solutions, homogeneous of degree k and translation invariant, of the generalized Laplace equation, since the latter are automatically constrained to be completely symmetrical; see Appendices B and C).

The result we have just proved implies that the (Gibbs) partition function for the 1-dimensional system composed of N identical oscillators coupled by pair harmonical potentials is the same (except for a multiplicative constant) for Bose and Fermi statistics [it is also the same (except for a multiplicative constant) for the system discussed in this paper, having additional pair inverse-square potentials of arbitrary strength (but the same for all pairs)[15]]. This fact does not appear to have been previously noted, possibly because the possibility to handle the many-oscillator problem in higher-dimensional spaces focused attention on the more realistic 3-dimensional case.[16]

3. GENERALIZED MODEL WITH QUADRATIC AND INVERSELY QUADRATIC POTENTIALS

In this section we discuss briefly the generalized problem characterized by the Hamiltonian

$$\begin{aligned} H = &- \frac{\hbar^2}{2m} \sum_{\alpha=1}^{A} \sum_{i=1}^{N_\alpha} \frac{\partial^2}{\partial x_{\alpha i}^2} \\ &+ \tfrac{1}{8} m\omega^2 \sum_{\alpha=1}^{A} \sum_{i=1}^{N_\alpha} \sum_{\beta=1}^{A} \sum_{j=1}^{N_\beta} (x_{\alpha i} - x_{\beta j})^2 \\ &+ \sum_{\alpha=1}^{A} g_\alpha \sum_{i=2}^{N_\alpha} \sum_{j=1}^{i-1} (x_{\alpha i} - x_{\alpha j})^{-2}. \end{aligned} \tag{3.1}$$

This Hamiltonian describes N particles, with

$$N = \sum_{\alpha=1}^{A} N_\alpha, \tag{3.2}$$

divided in A families of N_α particles, $\alpha = 1, 2, \cdots, A$. All the N particles interact pairwise through the harmonical potential $\frac{1}{4}m\omega^2(x_{\alpha i} - x_{\beta j})^2$, and in addition within each family (but not between different families) the particles interact pairwise through the potential $g_\alpha(x_{\alpha i} - x_{\alpha j})^{-2}$. To avoid 2-body collapse,[4,9] we assume as usual that

$$g_\alpha > -\hbar^2/(4m), \quad \alpha = 1, 2, \cdots, A. \tag{3.3}$$

The case of the preceding section corresponds to the present one with $A = 1$. The proofs of the results reported in this section parallel closely those of the corresponding results of Sec. 2, and are therefore omitted. We assume, whenever relevant, that within each family the particles are identical, satisfying Bose or Fermi statistics.

As in the preceding case, we need solve the eigenvalue problem only in the sector of configuration space characterized by the prescriptions

$$x_{\alpha i} \geq x_{\alpha, i+1},$$
$$\alpha = 1, 2, \cdots, A'; \; i = 1, 2, \cdots, N_\alpha - 1, \tag{3.4}$$

where A' is characterized by the condition

$$g_\alpha \neq 0, \quad \alpha = 1, 2, \cdots, A',$$
$$g_\alpha = 0, \quad \alpha = A' + 1, A' + 2, \cdots, A;$$

A' may be zero or it may coincide with A. Once the problem is solved within the sector (3.4), the extension of the many-body wavefunctions to the whole configuration space is easily achieved by the obvious generalization of the prescriptions discussed in detail at the beginning of the preceding section.

It is convenient to introduce the quantities

$$z_\alpha = \prod_{i=2}^{N_\alpha} \prod_{j=1}^{i-1} (x_{\alpha i} - x_{\alpha j}), \tag{3.5}$$

$$r^2 = \frac{1}{2N} \sum_{\alpha=1}^{A} \sum_{i=1}^{N_\alpha} \sum_{\beta=1}^{A} \sum_{j=1}^{N_\beta} (x_{\alpha i} - x_{\beta j})^2, \tag{3.6}$$

$$a_\alpha = \tfrac{1}{2}(1 + 4mg_\alpha\hbar^{-2})^{\frac{1}{2}}. \tag{3.7}$$

If $g_\alpha = 0$, $a_\alpha = \frac{1}{2}$ for Fermi statistics, $a_\alpha = -\frac{1}{2}$ for Bose statistics.

It can then be shown that the eigenfunctions of the eigenvalue equation

$$H\psi_{nk} = E_{nk}\psi_{nk} \tag{3.8}$$

have the form

$$\psi_{nk} = \left\{\prod_{\alpha=1}^{A} [z_\alpha^{a_\alpha + \frac{1}{2}}]\right\} \exp\left(-\frac{1}{2}\frac{m\omega}{\hbar}(\tfrac{1}{2}N)^{\frac{1}{2}}r^2\right) \times L_n^b\left[\frac{m\omega}{\hbar}(\tfrac{1}{2}N)^{\frac{1}{2}}r^2\right] \cdot P_k(x), \tag{3.9}$$

where L_n^b is a Laguerre polynomial,[12]

$$b = k + \tfrac{1}{2}(N - 3) + \sum_{\alpha=1}^{A} \tfrac{1}{2}N_\alpha(N_\alpha - 1)(a_\alpha + \tfrac{1}{2}), \tag{3.10}$$

and $P_k(x)$ is a translation-invariant polynomial in the N variables $x_{\alpha i}$, homogeneous of degree k and satisfying the generalized Laplace equation

$$\left[\sum_{\alpha=1}^{A} \sum_{i=1}^{N_\alpha} \frac{\partial^2}{\partial x_{\alpha i}^2} + 2\sum_{\alpha=1}^{A}(a_\alpha + \tfrac{1}{2}) \sum_{i=2}^{N_\alpha} \sum_{j=1}^{i-1} (x_{\alpha i} - x_{\alpha j})^{-1} \times \left(\frac{\partial}{\partial x_{\alpha i}} - \frac{\partial}{\partial x_{\alpha j}}\right)\right] P_k(x) = 0. \tag{3.11}$$

It can be shown that the polynomials $P_k(x)$ are completely symmetrical (invariant) for any coordinate exchange within the same family [provided the corresponding g_α does not vanish, or, if it vanishes, provided the particles satisfy Fermi statistics, so that $a_\alpha = \frac{1}{2}$; if, instead, g_α vanishes and the particles of the αth family satisfy Bose statistics, so that $a_\alpha = -\frac{1}{2}$, only solutions that are completely symmetrical in the variables $x_{\alpha i}$, $i = 1, 2, \cdots, N_\alpha$, should be considered, even though nonsymmetrical solutions of the generalized Laplace equation (3.11) exist].

The energy eigenvalues are given by the formula

$$E_{nk} = E_{2n+k} = \hbar\omega(\tfrac{1}{2}N)^{\frac{1}{2}}\left[\tfrac{1}{2}(N - 1) + \sum_{\alpha=1}^{A} \tfrac{1}{2}N_\alpha(N_\alpha - 1)(a_\alpha + \tfrac{1}{2}) + 2n + k\right]. \tag{3.12}$$

All these equations are straightforward generalizations of the corresponding equations of the preceding section. The analysis of the degeneracy of these energy levels could also be carried out in analogy to the treatment given there and in Appendix C.

4. THE SYSTEM WITH INVERSELY QUADRATIC POTENTIALS

The Hamiltonian of the system under consideration is

$$H = -\frac{\hbar^2}{2m}\sum_{i=1}^{N}\frac{\partial^2}{\partial x_i^2} + g\sum_{i=2}^{N}\sum_{j=1}^{i-1}(x_i - x_j)^{-2}, \tag{4.1}$$

again with the condition (2.2) to prevent 2-body collapse.[4,9] This Hamiltonian has of course no discrete spectrum; it describes only scattering states. The treatment of Sec. 2 and Appendix A implies that the complete set of stationary eigenfunctions of this

problem is (in the center-of-mass frame)

$$\psi_{pk} = z^{a+\frac{1}{2}} r^{-b} J_b(pr) P_k(x), \quad k = 0, 1, 2, \cdots, \quad p \geq 0, \tag{4.2}$$

with z, r, a, b, and $P_k(x)$ defined as in Sec. 2 [Eqs. (2.7)–(2.10) and (2.13)], and with p connected to the energy eigenvalue by

$$E = \hbar^2 p^2/(2m). \tag{4.3}$$

The product of r^{-b} times the Bessel function $J_b(pr)$ is the (regular) solution of the differential equation

$$-\frac{\hbar^2}{2m}\left[\frac{d^2}{dr^2} + (1+2b)\frac{1}{r}\frac{d}{dr} + p^2\right]\varphi(r) = 0, \tag{4.4}$$

which coincides with Eq. (2.11) when $\omega = 0$ and the energy E is given by Eq. (4.3). Of course, for each value of k there are $g(N, k)$ independent eigenfunctions, corresponding to the $g(N, k)$ independent solutions $P_k(x)$ of Eq. (2.10). We shall indicate them hereafter as $P_{kq}(x)$, using the quantum number q, which takes all integral values from 1 to $g(N, k)$, to label them. It is important to recall that, as proved in Appendix B, all these polynomials $P_{kq}(x)$ are completely symmetrical (invariant) under the exchange of any two coordinates x_i, x_j.

To simplify the discussion we assume hereafter that the particles are distinguishable. Attention need be confined only to the sector of phase space corresponding to a definite ordering of the particles, say that specified by the inequalities (2.4), whose validity is assumed hereafter.

The most general stationary eigenfunction of the Hamiltonian (4.1), corresponding to the eigenvalue (4.3), can be written in the form

$$\psi = z^{a+\frac{1}{2}} \sum_{k=0}^{\infty} \sum_{q=1}^{g(N,k)} c_{kq} r^{-A-k} J_{A+k}(pr) P_{kq}(x), \tag{4.5}$$

where to expose the dependence on the quantum number k we have introduced the convenient constant

$$A = b - k = \tfrac{1}{2}(N-3) + \tfrac{1}{2}N(N-1)(a+\tfrac{1}{2}). \tag{4.6}$$

To discuss scattering, we need only the asymptotic behavior of this function when all particles are far apart from each other. Then

$$\psi \sim \psi_{\text{in}} + \psi_{\text{out}}, \tag{4.7}$$

where

$$\psi_{\text{in}} \sim (\tfrac{1}{2}\pi pr)^{-\frac{1}{2}} z^{a+\frac{1}{2}} r^{-A} \times \sum_{k=0}^{\infty} \sum_{q=1}^{g(N,k)} c_{kq} r^{-k} e^{i(A+k+\frac{1}{2})\frac{1}{2}\pi - ipr} P_{kq}(x), \tag{4.8}$$

and

$$\psi_{\text{out}} \sim (\tfrac{1}{2}\pi pr)^{-\frac{1}{2}} z^{a+\frac{1}{2}} r^{-A} \times \sum_{k=0}^{\infty} \sum_{q=1}^{g(N,k)} c_{kq} r^{-k} e^{-i(A+k+\frac{1}{2})\frac{1}{2}\pi + ipr} P_{kq}(x). \tag{4.9}$$

The wavy symbol $\sim$ in these equations and below indicates asymptotic equality, i.e., equality up to corrections of order r^{-2}.

The stationary eigenfunction describing, in the center-of-mass frame, the scattering situation is characterized by the form

$$\psi_{\text{in}} \sim c \exp\left[i \sum_{i=1}^{N} p_i x_i\right], \tag{4.10}$$

with

$$p_i \leq p_{i+1}, \quad i = 1, 2, \cdots, N-1, \tag{4.11}$$

$$p^2 = \sum_{i=1}^{N} p_i^2, \tag{4.12}$$

and

$$\sum_{i=1}^{N} p_i = 0. \tag{4.13}$$

We now prove that if the constants c_{kq} are chosen so that Eq. (4.8) yields (4.10), then from Eq. (4.9) there also follows

$$\psi_{\text{out}} \sim e^{-iA\pi} c \exp\left[i \sum_{i=1}^{N} p_{N+1-i} x_i\right]. \tag{4.14}$$

In fact, the symmetry and the homogeneity of $P_{kq}(x)$ imply that

$$P_{kq}(-Sx) = e^{-ik\pi} P_{kq}(x), \tag{4.15}$$

where S indicates an arbitrary permutation of the coordinates x_i. We focus attention hereafter on the special permutation T, defined by

$$Tx_i = x_{N+1-i}, \quad i = 1, 2, \cdots, N. \tag{4.16}$$

The distinguishing property of this permutation is that the set $\{-Tx\}$ belongs to the same sector (2.4) of configuration space as the set $\{x\}$, because the inequalities

$$x_i \geq x_{i+1}, \quad i = 1, 2, \cdots, N-1, \tag{2.4}$$

also imply

$$-x_{N+1-i} \geq -x_{N-i}, \quad i = 1, 2, \cdots, N-1. \tag{4.17}$$

Now using Eq. (4.15) we can rewrite ψ_{out}, Eq. (4.9), in the form

$$\psi_{\text{out}} \sim e^{-i\pi A} (\tfrac{1}{2}\pi \bar{p} r)^{-\frac{1}{2}} z^{a+\frac{1}{2}} r^{-A} \times \sum_{k=0}^{\infty} \sum_{q=1}^{g(N,k)} c_{kq} r^{-k} e^{i(A+k+\frac{1}{2})\frac{1}{2}\pi - i\bar{p}r} P_{kq}(-Tx), \tag{4.18}$$

where we have set formally

$$\bar{p} = -p. \tag{4.19}$$

Let us re-emphasize that this representation of ψ_{out} applies in the sector (2.4) of configuration space. It

may therefore be compared with the representation of ψ_{in}, Eq. (4.8); the comparison yields, through Eq. (4.10), the result

$$\psi_{\text{out}} \sim e^{-i\pi A} c \exp\left[-i \sum_{i=1}^{N} \bar{p}_i T x_i\right], \qquad (4.20)$$

where $\bar{p}_i$ is related to $\bar{p}$ in the same manner as p_i is related to p. Since by dimensional arguments we can assert that

$$p_i = p\alpha_i, \qquad (4.21)$$

where the "angular" variables α_i are independent of p, Eq. (4.19) implies simply

$$\bar{p}_i = -p_i. \qquad (4.22)$$

Substituting this equality and Eq. (4.16) in Eq. (4.20) and changing the dummy index i into $N + 1 - i$ yield Eq. (4.14). QED

Let us now discuss the implications of the result we have just proved. The initial wavefunction (4.10) describes, in the sector (2.4) of configuration space, a (free) state where particle 1 has momentum p_1, particle 2 has momentum p_2, etc., the inequality (4.11) insuring that this is indeed an incoming scattering state, i.e., one where each particle gets less close to every other particle if time runs backward. The final wavefunction (4.14) describes, in the same sector (2.4) of configuration space, a (free) state where particle 1 has momentum p_N, particle 2 has momentum p_{N-1}, etc., the inequality (4.11) insuring that this is indeed an outgoing scattering state, i.e., one where each particle gets farther away from every other particle as time goes on. The result just proven implies that the stationary eigenfunction of the Hamiltonian H, Eq. (4.1), which is identified by the condition that its incoming part coincide with Eq. (4.10), contains only the outgoing wave Eq. (4.14); thus the initial state characterized by particle 1 having momentum p_1, particle 2 having momentum p_2, etc., can go only into the final state characterized by particle 1 having momentum p_N, particle 2 having momentum p_{N-1}, etc. This corresponds to the rule

$$p'_i = p_{N+1-i}, \quad i = 1, 2, \cdots, N, \qquad (4.23)$$

where p_i is the initial momentum of the ith particle and p'_i is the (only allowed value of the) final momentum of the same particle; we already noted above that the inequalities (4.11), satisfied by the initial momenta p_i, imply automatically that the final momenta p'_i given by Eq. (4.23) satisfy the inequalities

$$p'_i \geq p'_{i+1}, \quad i = 1, 2, \cdots, N-1, \qquad (4.24)$$

which characterize outgoing states. It should perhaps be emphasized that the rule (4.23) is not implied, for $N > 2$, by energy and momentum conservation [Eqs. (4.12) and (4.13), and the corresponding equations with p'_i in place of p_i], although it is of course consistent with these requirements.

The result (4.23) is certainly true, but nontrivial, also in the classical case; note that it is independent of the value of the coupling constant g, that only enters, through Eq. (4.6), in the phase factor $\exp(-i\pi A)$ multiplying Eq. (4.14).

The result (4.23) was already known for the case of infinitely repulsive zero-range δ-function interactions, in which case it is actually quite trivial in the classical case, although a bit less so in the quantal case[2]; that problem corresponds indeed to the limit of the present one as $g \to 0$, $a \to \frac{1}{2}$, since for extremely small g the interaction $g(x_i - x_j)^{-2}$ is effective only at very short interparticle separation $x_i \approx x_j$, where its singular nature continues to prevent the particles from crossing over—namely, it has the same effect as an infinitely repulsive δ-function potential.

Finally we discuss whether the result (4.23) holds only for the case of equal particles, or if it remains true even if the coupling constants g_{ij} multiplying the inversely quadratic potentials acting between the ith and jth particles are not all equal. The comparison with the infinitely repulsive zero-range δ-function case, as discussed above, suggests that this is, at least approximately, the case if all the constants g_{ij} are extremely small, independently from their being equal or different. But the measure of the smallness of the (nondimensionless) coupling constants g_{ij} is nontrivial. In the quantal case the dimensionless constants $2mg_{ij}\hbar^{-2}$ could be considered; but they become infinite in the classical limit, whereas the phenomenon just discussed should continue to be relevant in this limit. On the other hand, in the classical case an examination of the example discussed in Ref. 8 and in Appendix D suggests that the distinguishing parameter between the "small g" and the "not-small g" cases is the dimensionless quantity

$$\eta = \max_{i=1,2,\cdots,N-1} \left(\min_{t} \{[x_i(t) - x_{i+1}(t)]/r(t)\} \right), \qquad (4.25)$$

where t is the time and $r(t)$ is connected to the coordinates $x_i(t)$ through Eq. (2.8). But this quantity does not depend only on the parameters m and g_{ij} of the system; it depends also on the initial data p_i and a_i, defined by

$$x_i(t) \underset{t \to -\infty}{\longrightarrow} p_i t/m + a_i + O(t^{-1}). \qquad (4.26)$$

Therefore, no analogous quantity exists in the quantal case, since, in this case, we cannot specify both p_i and a_i. One might therefore be inclined to conjecture that the result (4.23) hold in all cases, for it should hold for small coupling constants g_{ij} independently from their being equal, and at the same time it should be independent of their scale. It is instead much more reasonable to conjecture that the result (4.23) does not hold, unless all the coupling constants (and the masses) are equal. In the quantal case, the asymptotic outgoing wavefunction ψ_{out}, corresponding to the asymptotic wavefunction ψ_{in} of Eq. (4.10), has, in general, the form

$$\psi_{\text{out}} \sim c \int dp_1' \cdots dp_N' \delta\left(\sum_{i=1}^{N} [p_i'^2 - p_i^2]\right) \delta\left(\sum_{i=1}^{N} [p_i' - p_i]\right) \times \left[\prod_{i=1}^{N-1} \theta(p_i' - p_{i+1}')\right] \exp\left[i \sum_{i=1}^{N} p_i' x_i\right] S(p_i, p_i'), \tag{4.27}$$

where the δ functions insure energy and momentum conservation, while the θ functions insure that the momenta satisfy the inequalities (4.24) characterizing a final scattering state. The S-matrix function, in the general case, need not contain any additional δ function and, in particular, need not reduce to a product of δ functions; only if all the coupling constants g_{ij} are equal (and all the masses are also equal), i.e., only in the case treated above, the function $S(p_i, p_i')$ (can be computed and) reduces to a product of δ functions[17]

$$S(p_i, p_i') = e^{-i\pi A}\left[\prod_{i=1}^{N} \delta(p_i' - p_{N+1-i})\right] \Big/ \left[\delta\left(\sum_{i=1}^{N} [p_i'^2 - p_i^2]\right) \delta\left(\sum_{i=1}^{N} [p_i' - p_i]\right)\right], \tag{4.28}$$

so that Eq. (4.27) reproduces Eq. (4.14). In addition, one conjectures that a measure of the deviation of the S matrix $S(p_i, p_i')$ for the general case from that of Eq. (4.28) be given by the quantity

$$\epsilon = \max_{i,j,i',j'} |(g_{ij} - g_{i'j'})| \Big/ \sum_{l=2}^{N} \sum_{k=1}^{l-1} |g_{lk}|, \tag{4.29}$$

which is clearly independent of the scale of the coupling constants g_{ij}. Of course, the value of this quantity is relevant to measuring the violation of the rule (4.23) also in the classical case, in addition to the quantity η of Eq. (4.25).

The above conjecture concerning the invalidity of the rule (4.23) in the general case (with different coupling constants g_{ij}) is strongly supported by the study of the 3-body classical case. Indeed in Appendix D it is proved that the rule (4.23) does not hold, at least in some cases, if the coupling constants g_{ij} are different; and the treatment suggests that if the coupling constants are different, the rule (4.23) almost never holds—namely, it is always violated except possibly for some symmetrical set of coupling constants and of initial data, the set of such initial data having, however, presumably a null measure relative to the whole set of possible initial conditions. These statements refer of course to the classical case; obviously they imply that, also in the quantal case, the S matrix does not reduce to the simple form (4.28), unless all the coupling constants, and the masses, be equal.

5. OPEN PROBLEMS

A number of questions are suggested by the results obtained. There is the problem of finding an explicit representation for the eigenfunctions, as it was done in the 3-body case.[4] This problem coincides with that of finding an explicit representation of the (translation-invariant) homogeneous polynomial solutions $P_k(x)$ of the generalized Laplace equations (2.10) and (3.11).[18]

Then there is the problem of solving the classical case, displaying explicitly the time evolution of the particle coordinates, for both types of models. For the first type (Secs. 2 and 3), it should be verified whether the plausible conjecture of periodic motion (already mentioned in the Introduction, and strongly suggested by the degenerate nature of the quantal spectrum) is confirmed.[7] For the second type (Sec. 4), the simple rule (4.23) connecting the initial momenta p_i to the final momenta p_i' should of course be recovered.

In view of the simplicity of the results characteristic of both types of models, it should be possible to reobtain the same conclusions, possibly more simply, by displaying the group-theoretical structure that certainly underlies them.[18] Note that this group-theoretical structure must be a peculiarity of the equal-particle models (see below).

As regards generalizations, the first problem that comes to mind is the extension to models characterized by different coupling constants g_{ij} of the inversely quadratic potential. Is such a generalization of the models of Secs. 2 and 3 still characterized by a completely[19] linear spectrum (in the quantal case, and by periodic motion, in the classical case)? It is plausible to conjecture that this is generally not the case [even though the linear formula

$$E_s = \hbar\omega(\tfrac{1}{2}N)^{\frac{1}{2}}\left[s + \tfrac{1}{2}(N-1) + \sum_{i=2}^{N} \sum_{j=1}^{i-1} (a_{ij} + \tfrac{1}{2})\right], \tag{5.1}$$

with

$$a_{ij} = \tfrac{1}{2}(1 + 4mg_{ij}\hbar^{-2})^{\frac{1}{2}}, \tag{5.2}$$

is an appealing generalization of Eq. (3.12), also in view of the simple physical interpretation it suggests, according to which the potential $g_{ij}(x_i - x_j)^{-2}$ produces simply an energy shift of the amount $\hbar\omega(\frac{1}{2}N)^{\frac{1}{2}}(a_{ij} \pm \frac{1}{2})$ relative to the case with $g_{ij} = 0$ and with the particles i and j being identical bosons (+ sign) or fermions (− sign)]. This conjecture is validated in Appendix E by a perturbative computation in the 3-body case.[20] For the second type of model (Sec. 4), we have already all but proved (in Sec. 4 and Appendix D) that the rule (4.23) breaks down unless all the coupling constants g_{ij} coincide.

A much more interesting generalization is of course in the direction of two or more space dimensions. It is expected that neither the completely linear nature of the spectrum for the first type of models nor the simple rule (4.23) for the second type of model is maintained, if the restriction to 1-space is dropped.

Another category of models that is somewhat similar to those considered in Secs. 2 and 3 of this paper and whose solution, even only in 1-dimensional space, would be most interesting, is defined by Hamiltonians such as

$$H = -\frac{\hbar^2}{2m}\sum_{i=1}^{N}\frac{\partial^2}{\partial x_i^2} + \tfrac{1}{4}m\omega^2\sum_{i=2}^{n}\sum_{j=1}^{i-1}(x_i - x_j)^2 + g\sum_{i=2}^{N}\sum_{j=1}^{i-1}(x_i - x_j)^{-2}, \tag{5.3}$$

with $n < N$. Already for $N = 3$ and $n = 2$ this Hamiltonian describes both a system with a discrete spectrum and one with scattering (between particle 3 and the bound state of particles 1 and 2, with the possibility of anelastic collisions occurring, although of course without break up). These two systems are characterized by (the same Hamiltonian but) different particle orderings, in the former case with particle 3 trapped between particles 1 and 2, in the latter case with particle 3 to the left (or to the right) of both particles 1 and 2.

Finally, we mention the possibility to interpret the results of this paper, and of the preceding ones,[4,8] in the context of the classical theory of wave propagation, say physical optics. In fact, these results imply the existence of a potential for which the fundamental equation of wave propagation

$$[\Delta + p^2 - V]\psi = 0 \tag{5.4}$$

can be solved exactly. Here, of course, Δ is the Laplace operator (in N dimensions; clearly the case $N = 3$ is the most interesting), and the potential is

$$V = G\sum_{i=2}^{N}\sum_{j=1}^{i-1}(x_i - x_j)^{-2}, \quad G > -\tfrac{1}{2}, \tag{5.5}$$

where now x_i are the (Cartesian) coordinates of a point in space.[21] This potential is of course not spherically symmetrical, and it is singular on the planes $x_i = x_j$; but it is otherwise acceptable within each of the wedges (sectors) in which the planes $x_i = x_j$ slice the whole space (in particular, it vanishes asymptotically). We leave to the interested reader the task of formulating in the language of physical optics the conclusions of Sec. 4 of this paper (or of Ref. 8). Such a translation is just the converse exercise to that performed by those (for instance, McGuire,[2] and before him Nussenzweig[2]) who invented solvable 1-dimensional problems by reinterpreting results that had been previously obtained in the framework of the theory of electromagnetic wave propagation.

ACKNOWLEDGMENTS

The main results of this paper have developed out of discussions with Yu. A. Simonov. It is a pleasure to acknowledge his contribution; it is regrettable that the pressure of other research commitments convinced him to opt out of co-authorship. I also wish to thank A. M. Perelomov for several illuminating discussions (mainly on the topics treated in Appendix C). The kind hospitality of the Theoretical Group at the Institute of Theoretical and Experimental Physics in Moscow during the academic year 1969–70, and the financial support of the Istituto Nazionale di Fisica Nucleare, are gratefully acknowledged.

APPENDIX A

In this appendix we report in detail the steps required to derive Eq. (2.11).

We begin by noting that

$$\sum_{i=1}^{N}\frac{\partial^2}{\partial x_i^2} = \frac{1}{N}\frac{\partial^2}{\partial y_0^2} + \sum_{i=1}^{N-1}\frac{\partial^2}{\partial y_i^2}, \tag{A1}$$

where the translation-invariant "Jacobi coordinates" are defined by[22]

$$y_i = [i(i+1)]^{-\frac{1}{2}}\left(ix_{i+1} - \sum_{j=1}^{i}x_j\right), \quad i = 1, 2, \cdots, N-1, \tag{A2}$$

while

$$y_0 = \frac{1}{N}\sum_{i=1}^{N}x_i \tag{A3}$$

is the center-of-mass coordinate. We also note that a transition from the "Cartesian" coordinates y_i, $i = 1, 2, \cdots, N-1$, to the "spherical" coordinates r, Ω_i, $i = 1, 2, \cdots, N-2$, with[23]

$$r^2 = \sum_{i=1}^{N-1} y_i^2, \tag{A4}$$

yields

$$\sum_{i=1}^{N} \frac{\partial^2}{\partial x_i^2} = \frac{\partial^2}{\partial y_0^2} + r^{2-N} \frac{\partial}{\partial r} r^{N-2} \frac{\partial}{\partial r} + r^{-2}\hat{L}, \tag{A5}$$

where the operator $\hat{L}$ acts only on the $N-2$ "angular" coordinates Ω_i.

It is important to note that *the radial coordinate r defined by Eq. (A4) coincides with that introduced previously through Eq. (2.8)*, as may be verified by an explicit computation.

Similarly, we note that

$$\sum_{i=2}^{N} \sum_{j=1}^{i-1} (x_i - x_j)^{-1} \left(\frac{\partial}{\partial x_i} - \frac{\partial}{\partial x_j} \right) = \tfrac{1}{2} N(N-1) \frac{1}{r} \frac{\partial}{\partial r} + r^{-2}\hat{M}, \tag{A6}$$

where the operator $\hat{M}$ acts only on the angular coordinates. This equation follows from the relation

$$\left(\frac{\partial}{\partial x_i} - \frac{\partial}{\partial x_j} \right) r = (x_i - x_j)/r, \tag{A7}$$

which is implied by the definition (2.8) of r.

Next we note that if $P_k(x)$ is a translation-invariant homogeneous polynomial of degree k in the N x_i coordinates, it is also a homogeneous polynomial of degree k in the $N-1$ y_i coordinates, and, therefore, $r^{-k}P_k(x)$ depends on the $N-2$ angular coordinates Ω_i but is independent of r. Using this remark, and Eqs. (A5) and (A6), we may cast Eq. (2.10) in the form

$$(\hat{L} + 2(a + \tfrac{1}{2})\hat{M}) r^{-k}P_k(x) = -k[k + N - 3 + N(N-1)(a + \tfrac{1}{2})] r^{-k}P_k(x). \tag{A8}$$

Finally, we note the two important relations

$$\sum_{i=1}^{N} \left(\frac{\partial}{\partial x_i} z^{a+\frac{1}{2}} \right) \frac{\partial}{\partial x_i} = (a + \tfrac{1}{2}) z^{a+\frac{1}{2}} \sum_{i=2}^{N} \sum_{j=1}^{i-1} (x_i - x_j)^{-1} \left(\frac{\partial}{\partial x_i} - \frac{\partial}{\partial x_j} \right), \tag{A9}$$

$$\sum_{i=1}^{N} \frac{\partial^2}{\partial x_i^2} z^{a+\frac{1}{2}} = 2mg\hbar^{-2} z^{a+\frac{1}{2}} \sum_{i=2}^{N} \sum_{j=1}^{i-1} (x_i - x_j)^{-2}. \tag{A10}$$

The first one follows directly from the definition of z, Eq. (2.7), which implies

$$\frac{1}{z} \frac{\partial z}{\partial x_i} = \sum_{\substack{j=1 \\ j \neq i}}^{N} (x_i - x_j)^{-1}, \tag{A11}$$

and from the identity

$$\sum_{i=1}^{N} \sum_{\substack{j=1 \\ j \neq i}}^{N} (x_i - x_j)^{-1} \frac{\partial}{\partial x_i} = \sum_{i=2}^{N} \sum_{j=1}^{i-1} (x_i - x_j)^{-1} \left(\frac{\partial}{\partial x_i} - \frac{\partial}{\partial x_j} \right). \tag{A12}$$

The second equation obtains evaluating the second derivative of $z^{a+\frac{1}{2}}$,

$$\frac{\partial^2}{\partial x_i^2} z^{a+\frac{1}{2}} = z^{a+\frac{1}{2}} \left[(a^2 - \tfrac{1}{4}) z^{-2} \left(\frac{\partial z}{\partial x_i} \right)^2 + (a + \tfrac{1}{2}) z^{-1} \frac{\partial^2 z}{\partial x_i^2} \right], \tag{A13}$$

and then using the definition of a, Eq. (2.9), and the two equations

$$z^{-2} \sum_{i=1}^{N} \left(\frac{\partial z}{\partial x_i} \right)^2 = 2 \sum_{i=2}^{N} \sum_{j=1}^{i-1} (x_i - x_j)^{-2}, \tag{A14}$$

$$\sum_{i=1}^{N} \frac{\partial^2 z}{\partial x_i^2} = 0, \tag{A15}$$

which are proved in C.

All the necessary tools having been prepared, we turn now to the derivation of Eq. (2.11). Substituting Eq. (2.6) into Eq. (2.3) and using Eqs. (A9) and (A10), we get

$$\left\{ -\frac{\hbar^2}{2m} \left[\sum_{i=1}^{N} \frac{\partial^2}{\partial x_i^2} + 2(a + \tfrac{1}{2}) \sum_{i=2}^{N} \sum_{j=1}^{i-1} (x_i - x_j)^{-1} \left(\frac{\partial}{\partial x_i} - \frac{\partial}{\partial x_j} \right) \right] + \tfrac{1}{4} m\omega^2 N r^2 - E \right\} r^{-k}P_k(x) r^k \varphi(r) = 0, \tag{A16}$$

where we have explicitly displayed the function $r^{-k}P_k(x)$, which depends on the $N-2$ angular variables Ω_i, and the function $r^k\varphi(r)$, which depends only upon the radial coordinate r. Then we use Eqs. (A5), (A6), and (A8), and we get

$$\left[-\frac{\hbar^2}{2m} \left(r^{2-N} \frac{d}{dr} r^{N-2} \frac{d}{dr} + N(N-1)(a + \tfrac{1}{2}) \frac{1}{r} \frac{d}{dr} - k[k + N - 3 + N(N-1)(a + \tfrac{1}{2})] r^{-2} \right) + \tfrac{1}{4} m\omega^2 N r^2 - E \right] r^k \varphi(r) = 0. \tag{A17}$$

This equation immediately yields Eq. (2.11). QED

APPENDIX B

In this appendix we prove that any polynomial solution of the generalized Laplace equation

$$L_G P(x) \equiv \left[\sum_{i=1}^{N} \frac{\partial^2}{\partial x_i^2} + G \sum_{i=2}^{N} \sum_{j=1}^{i-1} (x_i - x_j)^{-1} \left(\frac{\partial}{\partial x_i} - \frac{\partial}{\partial x_j}\right)\right] P(x) = 0 \tag{B1}$$

is completely symmetrical under the exchange of any two coordinates x_i, x_j. The only assumption needed for the proof is that

$$G \neq -2p, \quad p = 0, 1, 2, \cdots. \tag{B2}$$

This condition is of course fulfilled in the case of Eq. (2.10), since $G = 2(a + \frac{1}{2})$ is in fact positive.

Let us assume, *per absurdum*, that $P(x)$ is not symmetrical under the exchange of two coordinates, say x_1 and x_2. Then $P(x)$ and $P_{12}P(x)$, where P_{12} is the operator that interchanges x_1 and x_2, are different, and the polynomial

$$Q(x) = (1 - P_{12})P(x) \tag{B3}$$

does not vanish identically. This polynomial is also a solution of Eq. (B1), since P_{12} commutes with the (completely symmetrical) operator L_G of Eq. (B1); moreover, it is antisymmetrical under the exchange of the coordinates x_1 and x_2, and therefore it can be written in the form

$$Q(x) = (x_1 - x_2)^{2p+1} R(x), \tag{B4}$$

where p is an integer not less than zero and $R(x)$ is a polynomial, symmetrical under the exchange of x_1 and x_2 and not identically vanishing for $x_1 = x_2$:

$$R(x)|_{x_1 = x_2} \neq 0. \tag{B5}$$

But now from Eq. (B4) there follows that, in the neighborhood of $x_1 = x_2$,

$$(x_1 - x_2)^{-1} \left(\frac{\partial}{\partial x_1} - \frac{\partial}{\partial x_2}\right) Q(x) = 2(2p + 1)(x_1 - x_2)^{2p-1} R(x) + O[(x_1 - x_2)^{2p}]. \tag{B6}$$

From Eq. (B4) we also get

$$\left(\frac{\partial^2}{\partial x_1^2} + \frac{\partial^2}{\partial x_2^2}\right) Q(x) = 4p(2p + 1)(x_1 - x_2)^{2p-1} R(x) + O[(x_1 - x_2)^{2p}]. \tag{B7}$$

Inserting $Q(x)$ in Eq. (B1) and noting that all other terms besides those just computed produce contributions of order $(x_1 - x_2)^{2p}$, we finally obtain

$$2(2p + 1)[2p + G]R(x)(x_1 - x_2)^{2p-1} + O[(x_1 - x_2)^{2p}] = 0. \tag{B8}$$

Dividing this equation by $(x_1 - x_2)^{2p-1}$ and then setting $x_1 = x_2$ yields

$$2(2p + 1)[2p + G]R(x)|_{x_1 = x_2} = 0. \tag{B9}$$

This equation, together with Eq. (B2), is in contradiction with Eq. (B5); the proof *per absurdum* is therefore accomplished.

APPENDIX C

In this appendix we discuss some results concerning the multiplicity of the energy levels.

We begin with a proof of Eq. (2.18). Let us assume that two eigenfunctions of the form[24]

$$\psi_{nk} = z^{a+\frac{1}{2}} \exp\left[-\frac{1}{2}\frac{m\omega}{\hbar}(\tfrac{1}{2}N)^{\frac{1}{2}} r^2\right] \times L_n^b\left[\frac{m\omega}{\hbar}(\tfrac{1}{2}N)^{\frac{1}{2}} r^2\right] P_k(x), \tag{C1}$$

say $\psi_{n_1 k_1}$ and $\psi_{n_2 k_2}$, are mutually orthogonal (and therefore linearly independent) unless $n_1 = n_2$ and $k_1 = k_2$; also, let us recall that by definition $g(N, k)$ is the number of linearly independent eigenfunctions corresponding to an assigned value of the quantum number k. Then the number $f(N, k)$ of linearly independent eigenfunctions corresponding to the quantum number $s = 2n + k$ is the sum of the numbers $g(N, k)$ over all allowed (i.e., nonnegative integral) values of n and k that are consistent with the relation $s = 2n + k$:

$$f(N, s) = \sum_{n,k} g(N, k)\delta_{s,2n+k}\theta(n)\theta(k). \tag{C2}$$

Performing the sum over k by means of the Kronecker delta yields Eq. (2.18). QED

There remains to prove the orthogonality of $\psi_{n_1 k_1}$ and $\psi_{n_2 k_2}$. This (rather obvious) property is most conveniently proved by rewriting ψ_{nk} in the form

$$\psi_{nk} = R_{nk}(r)\chi_k(\Omega_i), \tag{C1'}$$

where

$$R_{nk}(r) = r^{b-\frac{1}{2}(N-3)} \exp\left[-\frac{1}{2}\frac{m\omega}{\hbar}(\tfrac{1}{2}N)^{\frac{1}{2}} r^2\right] \times L_n^b\left[\frac{m\omega}{\hbar}(\tfrac{1}{2}N)^{\frac{1}{2}} r^2\right], \tag{C3a}$$

$$\chi_k(\Omega_i) = [z/r^{\frac{1}{2}N(N-1)}]^{a+\frac{1}{2}} r^{-k} P_k(x), \tag{C3b}$$

and by noticing that the Hamiltonian operator H of Eq. (2.1) may be rewritten in the separated form

$$H = H_{CM} + H_r + r^{-2}H_\Omega, \tag{C4}$$

with

$$H_{CM} = -\frac{\hbar^2}{2mN}\frac{\partial^2}{\partial y_0^2}, \tag{C5}$$

$$H_r = -\frac{\hbar^2}{2m} r^{2-N}\frac{\partial}{\partial r} r^{N-2}\frac{\partial}{\partial r} + \tfrac{1}{4}Nm\omega^2 r^2, \tag{C6}$$

$$H_\Omega = -\frac{\hbar^2}{2m}\hat{L} + g\sum_{i=2}^{N}\sum_{j=1}^{i-1}[r/(x_i - x_j)]^2. \tag{C7}$$

Here we are using the notation, and some of the results, introduced at the beginning of Appendix A.

Using the separated form of the eigenfunction, Eq. (C1'), and of the Hamiltonian, Eq. (C4), we can recast the eigenvalue equation

$$H\psi_{nk} = E_{nk}\psi_{nk} \tag{C8}$$

into the separated form

$$H_\Omega\chi_k(\Omega_i) = \lambda_k\chi_k(\Omega_i), \tag{C9}$$

$$(H_r + r^{-2}\lambda_k)R_{nk}(r) = E_{nk}R_{nk}(r), \tag{C10}$$

and from Eqs. (C3a) and (C6) we get

$$\lambda_k = [\hbar^2/(2m)][b^2 - \tfrac{1}{4}(N-3)^2] \tag{C11a}$$

$$= [\hbar^2/(2m)][k + \tfrac{1}{2}N(N-1)(a+\tfrac{1}{2})] \times [k + N - 3 + \tfrac{1}{2}N(N-1)(a+\tfrac{1}{2})]. \tag{C11b}$$

Since the operators H_Ω and H_r are Hermitian, Eqs. (C9) and (C11) imply the orthogonality of $\chi_{k_1}(\Omega_i)$ and $\chi_{k_2}(\Omega_i)$ unless $k_1 = k_2$, and Eqs. (C10) and (2.14) imply the orthogonality of R_{n_1k} and R_{n_2k} unless $n_1 = n_2$. The combination of these two results implies the orthogonality of $\psi_{n_1k_1}$ and $\psi_{n_2k_2}$ unless $n_1 = n_2$ and $k_1 = k_2$. QED

[It can be explicitly verified, using the formula

$$\prod_{i=1}^{N} dx_i\,\delta\left(\sum_{i=1}^{N} x_i\right) \propto d\Omega r^{N-2}\,dr, \tag{C12}$$

where $d\Omega$ indicates the differential element for the "angular" coordinates Ω_i, that the orthogonality of R_{n_1k} to R_{n_2k} reproduce the well-known orthogonality relation for Laguerre polynomials.]

We turn now to a discussion of the multiplicity indices $g(N,k)$ and $f(N,k)$, beginning with some definitions. Let $a(N,k)$ be the number of linearly independent (completely symmetrical) polynomials of N variables x_i that are homogeneous of degree k. Let $b(N,k)$ be the number of linearly independent completely symmetrical polynomials of N variables x_i that are homogeneous of degree k and translation invariant [we show presently that $b(N,k)$ coincides with $f(N,k)$]. Let $c(N,k)$ be the number of linearly independent (completely symmetrical) polynomials of N variables x_i that are homogeneous of degree k and that are solutions of a given generalized Laplace equation such as Eq. (2.10). Finally, as in Sec. 2, let $g(N,k)$ be the number of linearly independent (completely symmetrical) polynomials that are homogeneous of degree k, that are translation invariant, and that are solutions of a given generalized Laplace equation such as Eq. (2.10). Then of course

$$b(N,k) = a(N,k) - a(N,k-1), \tag{C13}$$

because the requirement that $P_k(x)$ be translation invariant is equivalent to the condition that the polynomial $\sum_{i=1}^{N}\partial P_k(x)/\partial x_i$ vanish identically, and it corresponds therefore to $a(N,k-1)$ equations, namely as many equations as the number of different monomials that make up the general completely symmetrical polynomial homogeneous of degree $k-1$. An analogous argument yields

$$c(N,k) = a(N,k) - a(N,k-2), \tag{C14}$$

since the application of the (generalized) Laplace operator to a completely symmetrical homogeneous polynomial of degree k yields a completely symmetrical homogeneous polynomial of degree $k-2$. Analogous arguments also yield

$$g(N,k) = b(N,k) - b(N,k-2), \tag{C15a}$$

$$g(N,k) = c(N,k) - c(N,k-1), \tag{C15b}$$

and then either one of these equations yields

$$g(N,k) = a(N,k) - a(N,k-1) - a(N,k-2) + a(N,k-3). \tag{C16}$$

On the other hand, Eqs. (C15a) and (2.18) immediately yield

$$f(N,k) = b(N,k). \tag{C17}$$

An explicit formula for the computation of $a(N,k)$ is

$$a(N,k) = \sum_{\{n_i\}}\delta_{k,\Sigma}\prod_{i=1}^{N}\theta(n_i - n_{i-1}). \tag{C18}$$

Here all symbols are defined as in Eq. (2.23); indeed the validity of this equation is a consequence of the possibility of writing in the form (2.23) the most general polynomial that is homogeneous of degree k and completely symmetrical, for Eq. (C18) is obtained by counting the number of coefficients $a_{\{n_i\}}$ that enter in Eq. (2.23). A general closed expression for this sum is not known; the computation of $a(N,k)$ from this formula is a tedious task already for $N = 3$. Explicit

expressions for $a(N, k)$ for N up to 5 and arbitrary k have been obtained by a different procedure.[25] It is based on the possibility of writing the most general completely symmetrical polynomial in the alternative form

$$\sum a_{\{n_l\}}\delta_{k,\Sigma_{l=1}^{N} l n_l} \prod_{l=1}^{N} [S_l^{n_l}\theta(n_l)], \tag{C19}$$

where

$$S_l = \sum_{i=1}^{N} x_i^l, \tag{C20}$$

and the sum extends over all sets of N nonnegative integers n_l that are consistent with the homogeneity condition enforced by the Kronecker δ; note that it can be similarly asserted that the most general completely symmetrical and translation invariant polynomial of N variables, that is homogeneous of degree k, can be written in the form

$$\sum a_{\{n_l\}}\delta_{k,\Sigma_{l=2}^{N} l n_l} \prod_{l=2}^{N} [T_l^{n_l}\theta(n_l)], \tag{C21}$$

where

$$T_l = \sum_{i=1}^{N} (x_i - N^{-1}S_1)^l \tag{C22}$$

(so that obviously $T_1 = 0$ and T_l is translation invariant). The number $a(N, k)$ of linearly independent completely symmetrical polynomials of N variables, homogeneous of degree k, coincides with the number of coefficients $a_{\{n_l\}}$ entering in Eq. (C19); thus, it is the number of different sets $\{n_l; l = 1, 2, \cdots, N\}$ of N nonnegative integers that are consistent with the equation

$$k = \sum_{l=1}^{N} l n_l. \tag{C23}$$

Similarly Eq. (C22) implies that the number $b(N, k)$ of linearly independent completely symmetrical polynomials of N variables, homogeneous of degree k and translation invariant, coincides with the number of different sets $\{n_l; l = 2, 3, \cdots, N\}$ of $N - 1$ nonnegative integers that are consistent with the equation

$$k = \sum_{l=2}^{N} l n_l. \tag{C24}$$

But we proved above that this number coincides with the multiplicity index $f(N, k)$. Thus the statement reported in Sec. 2, and used there to obtain Eq. (2.22), is now proved.

The following trick[25] is convenient to evaluate $a(N, k)$. Introduce the generating function

$$A_N(z) = \sum_{k=0}^{\infty} a(N, k) z^k. \tag{C25}$$

The statement just proved implies the formula

$$A_N(z) = \sum_{n_1, n_2, \cdots, n_N = 0}^{\infty} z^{\Sigma_{l=1}^{N} l n_l}. \tag{C26}$$

From this we get

$$A_N(z) = \prod_{l=1}^{N} (1 - z^l)^{-1}. \tag{C27}$$

Separating this expression into partial fractions, and then re-expanding in powers of z and identifying the coefficients with those in Eq. (C25), yield explicit expressions for $a(N, k)$. In this manner, in Ref. 25, the following formulas were obtained:

$$a(2, k) = 1 + [\tfrac{1}{2}k], \tag{C28a}$$

$$a(3, k) = \{\tfrac{1}{12}(k + 2)(k + 4)\}, \tag{C28b}$$

$$a(4, k) = \{\tfrac{1}{144}(k + 2)(k^2 + 13k + 37 + \tfrac{9}{2}(1 + (-)^k))\}, \tag{C28c}$$

$$a(5, k) = [(\tfrac{1}{2880}((k + 1)(k + 2)(k + 3)(k + 24) + 155k^2 + 15k(67 + 3(-)^k))]. \tag{C28d}$$

In these four equations $[x]$ and $\{x\}$ indicate, respectively, the integral part of x and the integer closest to x. From these quantities and the formula

$$f(N, k) = a(n, k) - a(N, k - 1), \tag{C29}$$

which follows from Eqs. (C13) and (C17), one may immediately obtain explicit expressions for $f(N, k)$ for N up to 5 and arbitrary k. Alternatively, one may compute $f(N, k)$ directly in a similar manner, starting from the generating function

$$F_N(z) = \sum_{k=0}^{\infty} f(N, k) z^k \tag{C30a}$$

$$= \prod_{l=2}^{N} (1 - z^l)^{-1}. \tag{C30b}$$

In fact, if $a(N, k)$ is not already known, it is more convenient to compute $f(N, k)$ directly in this manner. Note that Eqs. (C30b) and (C27) imply

$$F_N(z) = (1 - z)A_N(z), \tag{C31}$$

and that this relation is consistent with Eqs. (C30a), (C25), and (C29). Also $g(N, k)$ is more easily obtainable directly from the generating function

$$G_N(z) = \sum_{k=0}^{\infty} g(N, k) z^k \tag{C32a}$$

$$= (1 - z^2)F_N(z) \tag{C32b}$$

$$= \prod_{l=3}^{N} (1 - z^l)^{-1}, \tag{C32c}$$

rather than from Eq. (C16) or

$$g(N, k) = f(N, k) - f(N, k - 2). \tag{C33}$$

Note that these expressions imply that $g(N, k)$ is the number of different solutions of the equation

$$k = \sum_{l=3}^{N} l n_l, \qquad , \quad \text{(C34)}$$

where n_l are nonnegative integers. Thus, for $N = 3$, $g(3, k)$ is unity if k is a multiple of 3 and vanishes otherwise, a result consistent with the findings of Ref. 4.

Finally we note that the generating function technique is also convenient to obtain the asymptotic behavior of the quantities $a(N, k)$, $f(N, k)$, and $g(N, k)$ at large k. This is done identifying the residue of the pole at $z = 1$,

$$\lim_{z\to 1} [(1 - z)^N A_N(z)] = (N!)^{-1}, \quad \text{(C35a)}$$

$$\lim_{z\to 1} [(1 - z)^{N-1} F_N(z)] = (N!)^{-1}, \quad \text{(C35b)}$$

$$\lim_{z\to 1} [(1 - z)^{N-2} G_N(z)] = 2/N!, \quad \text{(C35c)}$$

with the singular behavior of the power expansions of these functions at $z = 1$. In this manner the asymptotic expression

$$\lim_{k\to\infty} [a(N, k) k^{1-N}] = [N!\,(N - 1)!]^{-1} \quad \text{(C36a)}$$

is obtained in Ref. 25. In an analogous manner one gets

$$\lim_{k\to\infty} [f(N, k) k^{2-N}] = [N!\,(N - 2)!]^{-1}, \quad \text{(C36b)}$$

$$\lim_{k\to\infty} [g(N, k) k^{3-N}] = 2/[N!\,(N - 3)!]. \quad \text{(C36c)}$$

APPENDIX D

In this appendix we study [in the sector (2.4)] the classical 1-dimensional 3-body problem with pair inversely quadratic potentials of unequal strength, and we produce an explicit example that violates the rule

$$p_i' = p_{4-i}, \quad \text{(D1)}$$

where p_i indicates the initial momentum of the ith particle and p_i' its final momentum. This rule is of course enforced if the coupling constants are all equal; an explicit proof in the classical case has been given by Marchioro,[8] whose treatment we follow here.

The Hamiltonian of the problem is

$$H = \frac{1}{2m} \sum_{i=1}^{3} p_i^2 + g_3(x_1 - x_2)^{-2} + g_1(x_2 - x_3)^{-2} + g_2(x_3 - x_1)^{-2}, \quad \text{(D2)}$$

where the coupling constants g_i are all nonnegative. It is convenient to introduce[4,8] the variables r and φ, defined by

$$r \cos\varphi = (x_1 + x_2 - 2x_3)/\sqrt{6}, \quad \text{(D3a)}$$

$$r \sin\varphi = (x_1 - x_2)/\sqrt{2}. \quad \text{(D3b)}$$

Then in the center-of-mass frame the Hamiltonian becomes

$$H = \frac{1}{2m}(p_r^2 + r^{-2} p_\varphi^2) + \tfrac{1}{2} r^{-2}\{g_3(\sin\varphi)^{-2} + g_1[\sin(\varphi + \tfrac{2}{3}\pi)]^{-2} + g_2[\sin(\varphi + \tfrac{4}{3}\pi)]^{-2}\}, \quad \text{(D4)}$$

where p_r and p_φ are the momenta conjugate to r and φ:

$$p_r = m \frac{dr}{dt}, \quad \text{(D5)}$$

$$p_\varphi = m r^2 \frac{d\varphi}{dt}. \quad \text{(D6)}$$

The separability of the Hamiltonian (D4) implies the existence of a second constant of the motion B^2, in addition to the energy E:

$$E = \frac{1}{2m} p_r^2 + B^2 r^{-2}, \quad \text{(D7)}$$

$$B^2 = \frac{1}{2m} p_\varphi^2 + \tfrac{1}{2}\{g_3(\sin\varphi)^{-2} + g_1[\sin(\varphi + \tfrac{2}{3}\pi)]^{-2} + g_2[\sin(\varphi + \tfrac{4}{3}\pi)]^{-2}\}. \quad \text{(D8)}$$

From Eqs. (D7) and (D5) we get[8]

$$r(t) = \left(\frac{2E}{m} t^2 + r_{\min}^2\right)^{\frac{1}{2}}, \quad \text{(D9)}$$

where

$$r_{\min} = r(0) = BE^{-\frac{1}{2}}; \quad \text{(D10)}$$

then from Eqs. (D8), (D6), and (D9) we get

$$F[\varphi(t)] = F[\varphi(-\infty)] + \operatorname{arctg}(t/T) + \tfrac{1}{2}\pi, \quad \text{(D11)}$$

where

$$T = r_{\min}(m/2E)^{\frac{1}{2}} = (B/E)(\tfrac{1}{2}m)^{\frac{1}{2}}, \quad \text{(D12)}$$

and

$$F(\varphi') = \int^{\varphi'} d\varphi [1 - \tfrac{1}{2} B^{-2}\{g_3[\sin\varphi]^{-2} + g_1[\sin(\varphi + \tfrac{2}{3}\pi)]^{-2} + g_2[\sin(\varphi + \tfrac{4}{3}\pi)]^{-2}\}]^{-\frac{1}{2}}. \quad \text{(D13)}$$

We are interested in the asymptotic behavior of the particles in the remote past and future, when they are widely separated and move freely. We set by definition

$$x_1(t) \xrightarrow[t\to+\infty]{} p_1' t/m + a_1' + O(t^{-1}), \quad \text{(D14)}$$

$$x_i(t) \xrightarrow[t\to-\infty]{} p_i t/m + a_i + O(t^{-1}), \quad \text{(D15)}$$

and since we are working in the center-of-mass frame, we also have

$$\sum_{i=1}^{3} p_i = \sum_{i=1}^{3} p_i' = 0. \tag{D16}$$

Moreover the initial momenta p_i must be consistent with the inequalities

$$p_1 \leq p_2 \leq p_3 \tag{D17a}$$

characterizing an incoming state (particles approaching each other), while the momenta p_i' shall be consistent with the opposite inequalities

$$p_1' \geq p_2' \geq p_3' \tag{D17b}$$

characterizing an outgoing state. For simplicity, we also choose the origin of the x axis to coincide with the center-of-mass, so that

$$\sum_{i=1}^{3} a_i = \sum_{i=1}^{3} a_i' = 0. \tag{D18}$$

Now from Eq. (D9) we get

$$r(t) \xrightarrow[t\to\pm\infty]{} \pm(2E/m)^{\frac{1}{2}}t + O(t^{-1}); \tag{D19}$$

from this equation and Eq. (D3b) we obtain

$$\varphi(-\infty) = \arcsin\,[\tfrac{1}{2}(p_2 - p_1)(p_1^2 + p_2^2 + p_1p_2)^{-\frac{1}{2}}]. \tag{D20}$$

We have also used the obvious relation

$$E = \frac{1}{2m}\sum_{i=1}^{3} p_i^2 = \frac{1}{m}(p_1^2 + p_2^2 + p_1p_2), \tag{D21}$$

which can be explicitly obtained from Eqs. (D15), (D16), and (D19), and from the relation

$$r^2 = \tfrac{1}{3}[(x_1 - x_2)^2 + (x_2 - x_3)^2 + (x_3 - x_1)^2], \tag{D22}$$

which is implied by Eqs. (D3).

But once $\varphi(-\infty)$ is known, $\varphi(+\infty)$ can be evaluated from the implicit equation

$$F[\varphi(+\infty)] = F[\varphi(-\infty)] + \pi, \tag{D23}$$

which is a special case of Eq. (D11); from $\varphi(+\infty)$ the final momenta are determined through

$$p_1' - p_2' = 2(p_1^2 + p_2^2 + p_1p_2)^{\frac{1}{2}} \sin\varphi(+\infty), \tag{D24}$$

$$p_1' + p_2' = \left(\frac{2}{\sqrt{3}}\right)(p_1^2 + p_2^2 + p_1p_2)^{\frac{1}{2}} \cos\varphi(+\infty), \tag{D25}$$

which follow from Eqs. (D14), (D19), and (D3). Note, however, that the implicit function $F[\varphi]$ contains also the constant B^2, which must also be determined from the initial data. This can be done from Eq. (D8), which yields

$$B^2 = \frac{1}{2m}p_\varphi^2(-\infty) + \tfrac{1}{2}\{g_3[\sin\varphi(-\infty)]^{-2} + g_1[\sin(\varphi(-\infty) + \tfrac{2}{3}\pi)]^{-2} + g_2[\sin(\varphi(-\infty) + \tfrac{4}{3}\pi)]^{-2}\}. \tag{D26}$$

As for $p_\varphi(-\infty)$, it can be obtained from Eqs. (D6), (D3), (D15), and (D18), and we find

$$p_\varphi(-\infty) = \sqrt{3}(p_1a_2 - p_2a_1). \tag{D27}$$

Note that $p_\varphi(-\infty)$, and therefore also B^2, depend not only on the initial momenta p_i, but also on the constants a_i of Eq. (D15).

All the steps for the computation of the final momenta p_i' from the initial data p_i and a_i are now ready, the relevant equations being (D16), (D20) [to evaluate $\varphi(-\infty)$ from the initial momenta p_i], (D16), (D18), (D27), (D26) (to evaluate the constant B^2 from the initial data p_i and a_i), (D23) [to evaluate $\varphi(+\infty)$ from $\varphi(-\infty)$ and B^2], and finally (D16), (D24), and (D25) [to evaluate the final momenta p_i' from $\varphi(+\infty)$]. The remaining difficulty is the inability to perform explicitly, in the general case, the integral in the definition of $F(\varphi)$, Eq. (D13). In the special case

$$g_1 = g_2, \tag{D28}$$

to which attention is hereafter confined, this integral can be cast into the form

$$F(\varphi) = \frac{-1}{2}\int^{\cos^2\varphi} dx(1 - 4x) \times (c_1x + c_2x^2 + c_3x^3 + c_4x^4)^{-\frac{1}{2}}, \tag{D29}$$

where

$$c_1 = 1 - \tfrac{1}{2}(8g_1 + g_3)B^{-2}, \tag{D30a}$$

$$c_2 = -9 + 4(g_3 - g_1)B^{-2}, \tag{D30b}$$

$$c_3 = 24 - 8(g_3 - g_1)B^{-2}, \tag{D30c}$$

$$c_4 = -16. \tag{D30d}$$

If in addition we assume the initial data to be such that

$$c_1 = 0 \tag{D31a}$$

or, equivalently,

$$B^2 = \tfrac{1}{2}(8g_1 + g_3), \tag{D31b}$$

then the integral of Eq. (D29) can be performed and it yields

$$F(\varphi) = \frac{-1}{2}\left((9 + \gamma)^{-\frac{1}{2}} \arcsin\frac{(12 + \gamma)\cos^2\varphi - 9 - \gamma}{\cos^2\varphi\,[\gamma(\gamma + 8)]^{\frac{1}{2}}} + \arcsin\frac{-16\cos^2\varphi + 12 + \gamma}{[\gamma(\gamma + 8)]^{\frac{1}{2}}}\right), \tag{D32}$$

with

$$\gamma = 4(g_1 - g_3)B^{-2} = 8(g_1 - g_3)/(8g_1 + g_3). \tag{D33}$$

We are assuming that γ is positive.

The derivation of $\varphi(+\infty)$ from $\varphi(-\infty)$ through Eqs. (D32) and (D23) still involves the solution of a transcendental equation; it cannot therefore be done explicitly. But the main purpose of this discussion is to prove that the rule (D1) need not hold if the three coupling constants g_i are not all equal. This is easily done by assuming its validity in one specific example, and showing that there results an inconsistency. Thus, we assume

$$p_1 = -p, \quad p_2 = 0, \quad p_3 = p, \tag{D34}$$

and

$$p_1' = p, \quad p_2' = 0, \quad p_3' = -p, \tag{D35}$$

with p positive [so that the inequalities (D17) are satisfied]. The choice (D34) implies, through Eq. (D20),

$$\varphi(-\infty) = \tfrac{1}{6}\pi, \tag{D36}$$

while the choice (D35) yields

$$\varphi(+\infty) = \tfrac{1}{6}\pi = \varphi(-\infty). \tag{D37}$$

[This is consistent with the remark[8] that generally the rule (D1) corresponds to $\varphi(\infty) = \frac{1}{3}\pi - \varphi(-\infty)$].

But the result (D37) is clearly inconsistent with Eqs. (D23) and (D32), at least so long as the arguments of the arcsin functions of Eq. (D32) are in the interval between -1 and 1. [Note that this is never the case if $g_1 = g_3$, i.e., if all coupling constants are equal; in fact, in this case, Eq. (D32) is meaningless, because γ vanishes. Indeed, the correct integration of Eq. (D29) in this case reproduces Marchioro's results,[8] yielding the rule (D1).] When g_1 is larger than g_3, this condition is easily satisfied, for instance, by the simple choice $g_3 = 0$, in which case $\gamma = 1$ and the arguments of the arcsin functions become $-\frac{1}{9}$ and $\frac{1}{3}$, respectively. As for the condition (D31), which we had to assume in order to perform explicitly the integration that led to Eq. (D32), it is easily seen that, through Eqs. (D27), (D34), and (D26), it becomes

$$a_2^2 = m(g_1 - g_3)/p^2,$$

so that it can always be enforced by appropriate choice of the initial constant a_2.

APPENDIX E

In this appendix we disprove the conjecture that the spectrum of the problem of Sec. 2 is completely linear even if the coupling constants of the inversely quadratic interaction are different for different pairs. This we do by displaying an explicit example that violates this conjecture. This is the 3-body model characterized by the Hamiltonian

$$H = H_0 + \epsilon(x_1 - x_2)^{-2}, \tag{E1}$$

where H_0 is the Hamiltonian of Sec. 2, Eq. (2.1), with $N = 3$, and ϵ is a small parameter. The spectrum of this model is then given by the formula[4]

$$E_{nl} = \hbar\omega(\tfrac{3}{2})^{\frac{1}{2}}[2n + B_l(\epsilon) + 1], \tag{E2}$$

where $B_l^2(\epsilon)$ are the eigenvalues of the differential equation[4]

$$\left(-\frac{\hbar^2}{2m}\frac{\partial^2}{\partial\varphi^2} + \frac{9g}{2\sin^2 3\varphi} + \frac{\epsilon}{2\sin^2\varphi}\right)F_l(\varphi) = B_l^2(\epsilon)F_l(\varphi), \quad 0 \le \varphi \le \tfrac{1}{3}\pi. \tag{E3}$$

For $\epsilon = 0$, the complete set of eigenfunctions of this equation are[4]

$$F_l(\varphi) = (\sin 3\varphi)^{a+\frac{1}{2}}C_l^{a+\frac{1}{2}}(\cos 3\varphi), \quad l = 0, 1, 2, \cdots, \tag{E4}$$

where C_l^p is a Gegenbauer polynomial,[12] and the corresponding eigenvalues are

$$B_l(0) = 3(l + a + \tfrac{1}{2})(\hbar^2/2m)^{\frac{1}{2}}, \quad l = 0, 1, 2, \cdots. \tag{E5}$$

Here, a is defined by Eq. (2.9).

To first order in ϵ the eigenvalues $B_l^2(\epsilon)$ are given by

$$B_l^2(\epsilon) = B_l^2(0) + \tfrac{1}{2}\epsilon\beta_l, \tag{E6}$$

where

$$\beta_l = \int_0^{\frac{1}{3}\pi} d\varphi \sin^{-2}\varphi F_l^2(\varphi) \Big/ \int_0^{\frac{1}{3}\pi} d\varphi F_l^2(\varphi). \tag{E7}$$

Thus, to first order in ϵ the energy spectrum is

$$E_{nl} = \hbar\omega(\tfrac{3}{2})^{\frac{1}{2}}[2n + 3l + 3a + \tfrac{5}{2} + \tfrac{1}{6}\epsilon\beta_l(l + a + \tfrac{1}{2})^{-1}m\hbar^{-2}]. \tag{E8}$$

Therefore, if the spectrum is to be completely linear, it should be true that

$$\beta_l = \mu(l + a + \tfrac{1}{2}) + \nu(l + a + \tfrac{1}{2})^2, \tag{E9}$$

where μ and ν are numerical constants. This implies the condition

$$6\beta_0 - 8\beta_1 + 3\beta_2 = 0. \tag{E10}$$

A simple example that demonstrates that this equation is not verified obtains considering the special case where $a = \frac{3}{2}$.[26] Then with a little algebra this condition becomes

$$\int_0^{\frac{1}{3}\pi} d\varphi \sin 3\varphi[\sin 15\varphi - \tfrac{5}{9}\sin 9\varphi][\sin 3\varphi/\sin\varphi]^2 = 0, \tag{E11}$$

and it is easily seen that this equality does not hold.

Note that if the perturbing potential had maintained the equality of the coupling constants, i.e., if it had

been of the form

$$\epsilon[(x_1 - x_2)^{-2} + (x_2 - x_3)^{-2} + (x_3 - x_1)^{-2}] = \tfrac{1}{2}\epsilon r^{-2} \sin^{-2} 3\varphi, \quad \text{(E12)}$$

the term $(\sin 3\varphi/\sin \varphi)^2$ in the integrand of Eq. (E11) would be missing; then of course this equation would be satisfied. Indeed, in such a case Eq. (E9) holds, with $\nu = 0$ and $\mu = 9/a$.

Although we have for simplicity displayed only a very special example in which the conjecture of complete linearity of the spectrum is violated, it is obvious that, for 1-dimensional models with quadratic and inversely-quadratic forces, this property is the exception rather than the rule, holding only in some special cases such as those considered in Sec. 2 (all particles equal) or in Sec. 3 (all particles equal within each family, the same quadratic interaction between all pairs, no inversely quadratic interaction between different families).

*Present address: Physics Dept., Rome University, Rome, Italy.

[1] E. H. Lieb and D. C. Mattis, *Mathematical Physics in One Dimension* (Academic, New York, 1966).

[2] F. A. Berezin, G. P. Pochil, and V. M. Finkelberg, Moscow Univ. Vestnik **1**, 21 (1964) (in Russian); J. B. McGuire, J. Math. Phys. **5**, 622 (1964); E. Brezin and J. Zinn-Justine, Compt. Rend. Acad. Sci. (Paris) **B263**, 670 (1966); C. N. Yang, Phys. Rev. Lett. **19**, 1312 (1967); Phys. Rev. **168**, 1920 (1968). The 1-dimensional 3-body problem with zero-range interactions had been previously solved in some special cases. For a detailed treatment of one such example, and a review of work in this field up to 1960, see H. M. Nussenzweig, Proc. Roy. Soc. (London) **A264**, 408 (1961).

[3] A complete analysis of the S matrix describing all scattering processes between these bound states has been given by Yang (see the second of his papers, Ref. 2 above).

[4] F. Calogero, J. Math. Phys. **10**, 2191 (1969).

[5] F. Calogero, J. Math. Phys. **10**, 2197 (1969), hereafter referred to as C. This paper is marred by several misprints: In the second term in Eq. (2.8), $\partial^2 z/\partial x_i^2$ should appear in place of $\partial z/\partial x_i$; in Eqs. (2.15) and (2.18), $\frac{1}{2}$ should appear in place of m within the square root; in Eq. (2.20), r^2 should appear in place of z^2 in the argument of the exponential; in Eq. (2.22), the plus sign after a should be a minus sign; in the sentence before Eq. (2.4), the word "symmetrical" should be replaced by "antisymmetrical and symmetrical."

[6] The coincidence of the energy levels (except for a constant shift) is essentially trivial, but the coincidence of the multiplicities (that probably is peculiar to the 1-dimensional case) is remarkable (see below).

[7] For $N = 3$, this has been proved by C. Marchioro (unpublished).

[8] C. Marchioro, J. Math. Phys. **11**, 2193 (1970).

[9] See, for instance, L. D. Landau and E. M. Lifshitz, *Quantum Mechanics* (Pergamon, New York, 1958), Sec. 35.

[10] Actually, the derivative of the wavefunction with respect to the variable x_i vanishes, when x_i coincide with another variable x_j, only if the inversely quadratic potential is repulsive ($g > 0$); but when multiplied with the wavefunction itself, it vanishes in all cases; and it is this that counts (see below and Ref. 4).

[11] The solutions considered in C correspond to the subset of these solutions characterized by the condition $k = 0$.

[12] I. S. Gradshteyn and I. M. Ryzhik, *Tables of Integrals, Series and Products* (Academic, New York, 1965); A. Erdélyi, Ed., *Higher Transcendental Functions* (McGraw-Hill, New York, 1953), Vol. II.

[13] Nonpolynomial solutions of Eq. (2.10) contain singularities and are therefore excluded by the requirement that the wavefunction (2.6) be physically acceptable (for an explicit illustration in the 3-body case, see Ref. 4). An additional argument to justify the absence of physically acceptable nonpolynomial solutions of Eq. (2.10) is that the polynomial solutions alone constitute a complete set [in the sector (2.4); see below].

[14] The ansatz (2.6), when inserted in the eigenvalue equation (2.3), yields Eqs. (2.10) and (2.11), provided $a^2 = \frac{1}{4}(1 + 4mg\hbar^{-2})$ (see Appendix A). The positive solution of this quadratic equation is the only acceptable one if g does not vanish, but both solutions $a = \pm\frac{1}{2}$ are acceptable (in the whole configuration space) if g vanishes. Acceptability is conditioned by the continuity of ψ^2 and $\psi\psi'$ at $x_i = x_j$, the prime indicating here differentiation relative to x_i or x_j.

[15] While the proportionality of the partition functions implies that these systems are thermodynamically equivalent, differences would show up in observables (such as, for instance, correlation functions) depending not only on the energy spectrum of the system but also on the form of its many-body wavefunction. Incidentally, the partition function is trivially computed from Eq. (2.26); indeed, up to a multiplicative constant, it coincides with the generating function $F_N(z)$ of Appendix C, with $z = \exp(\hbar\omega(\frac{1}{2}N)^{\frac{1}{2}}/KT)$. The (more interesting) problem of the gas composed of N 1-dimensional particles interacting by inversely square potentials has been studied, in the $N \to \infty$ limit with constant density, by C. Marchioro and E. Presutti, Nuovo Cim. Lett. **4**, 488 (1970); and by Bill Sutherland, J. Math. Phys. **11**, 3183 (1970).

[16] R. G. Storer, Phys. Rev. Lett. **24**, 5 (1970).

[17] We use for simplicity a symbolic notation, with δ functions in the denominator, whose significance should be self-evident [see Eq. (4.27)].

[18] Some progress in this direction has been made by A. M. Perelomov (private communication).

[19] Of course the separability of the Hamiltonian into "radial" and "angular" parts (see Appendices C and A), which obtains independently of the equality of the particles, and the simple nature of the harmonical potential, imply that the spectrum depends in all cases linearly upon the "radial" quantum number n.

[20] I owe to A. M. Perelomov the suggestion to test this conjecture by perturbation theory.

[21] Other, possibly more interesting, potentials, for which the basic equation of wave propagation can be solved exactly, are obtained from the above one by going over to the "Jacobi" coordinates (and eliminating the "center-of-mass coordinate" y_0 from the Laplace operator, thereby reducing by one the dimensionality of the space it refers to; see Appendix A).

[22] Different definitions of the Jacobi coordinates could be chosen without affecting the final result. See, e.g., M. Grynberg and Z. Koba, Ann. Phys. (N.Y.) **26**, 418 (1964).

[23] The exact choice of the angular coordinates Ω_i is immaterial, the only important property being that of Eq. (A5) below. A possible choice are the standard polar coordinates; another possible choice are the so called zonal coordinates. See, e.g., P. Appel and J. Kampé de Fériet, *Fonctions Hypergéométriques et Hyperspheriques* (Gauthier-Villars, Paris, 1926), Pt. II, Chap. II.

[24] Here and below z, r, a, and b are defined as in Sec. 2, Eqs. (2.7)–(2.9) and (2.13), and $P_k(x)$ is a translation-invariant polynomial in the N variables x_i, homogeneous of degree k, and satisfying the generalized Laplace equation (2.10).

[25] A. M. Perelomov, V. S. Popov, and I. A. Malkin, ITEF preprint 337 [a slightly abridged version of this paper, in which the results reported here have been omitted, has been published: Sov. J. Nucl. Phys. **2**, 533 (1965)].

[26] The first three Gegenbauer polynomials with superscript 2 are

$$C_0^2(x) = 1, \quad C_1^2(x) = x, \quad C_2^2(x) = 1 - 6x^2.$$

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 12, NUMBER 3 MARCH 1971

Structure of the Combinatorial Generalization of Hypergeometric Functions for $SU(n)$ States*

ALFRED C. T. WU
Department of Physics, University of Michigan, Ann Arbor, Michigan 48104

(Received 19 August 1970)

The combinatorics of the boson operator formalism in the construction of the $SU(n)$ states provides a natural scheme for the appearance of certain generalized hypergeometric functions. It is shown that, while special cases exist where the functions thus generated belong to the class of generalized hypergeometric functions defined by Gel'fand *et al.* as being the Radon transforms of products of *linear* forms, the general cases apparently do not. This is already so at the $SU(4)$ level.

1. INTRODUCTION

In the representation theory of unitary groups, much has been known on the problem of explicit construction of state vectors for the irreducible representation, in particular, with the device of the boson operator formalism.[1] One of the crucial problems lies in the structural analysis of a general state vector such that hopefully the algebraic complexities which mount extremely rapidly as the rank of the group goes up may be systematically controlled.

In recent years, some efforts have been made in, among other things, the combinatorial structure of the state vectors for the irreducible representation of $U(n)$.[2-5] In particular, it was found[2-5] that one can readily obtain the normalization constants by carrying out essentially the following steps: (i) A general state vector is obtained by applying an appropriate string of lowering operators from a so-called semimaximal state,[6] the latter being expressed as appropriate products of (antisymmetrized) determinantal factors of creation operators (each factor being raised to an appropriate power) acting on the vacuum state; (ii) the process of pushing through the lowering operators all the way to the right results in the combinatorics as a consequence of the canonical commutation relations.

One of the problems then is to study the combinatorial structure entailed in the step (ii) above. More explicitly, Baird and Biedenharn[2] have shown that for $SU(3)$ the operator-valued polynomials can, in fact, be formally expressed in terms of the well-known Gauss hypergeometric function ${}_2F_1(a, b; c; x)$; namely,

$$|\text{general } SU(3) \text{ state}\rangle = \text{const (product of antisymmetrized creation operators) } {}_2F_1(a, b; c; x)\,|0\rangle, \quad (1)$$

where the coefficients a, b, and c depend linearly on the Gel'fand labels[7] of the state, while the variable x is formally an operator quotient in such a way that all the denominators are cancelled eventually by appropriate multiplicative factors outside the hypergeometric function.

Attempts have been made by Ciftan and Biedenharn[4] and in particular by Ciftan[5] to generalize Eq. (1) to higher rank unitary groups.[8] Unfortunately, because of the increasing algebraic complexities, no simple expressions analogous to Eq. (1) are known. For $SU(4)$, Ciftan was able to recognize the structure of the individual block constituents, where each block of terms essentially corresponds to the action of one particular lowering operator (raised to a power), but the expression for a general $SU(4)$ state was left as a sixfold sum of operators.

The main purpose of this paper is to answer some questions raised by Ciftan's treatment of the $SU(4)$ state. The way that combinatorics of boson calculus provides a natural scheme for the appearance of certain generalized hypergeometric functions is rather intriguing. In view of considerable interest in the connection between special functions and group theory,[9-12] it is perhaps desirable to ascertain what class of functions does such generalization via combinatorics lead to.

In fact, at the $SU(n)$ level, the Gauss function ${}_2F_1$ appears as a consequence of the action of the lowering operator $(L_{n-1}^{n-2})^{\alpha}$. The Appell function[13,14] of the second kind, F_2, is obtained from the action of $(L_{n-1}^{n-3})^{\alpha}$. The Lauricella function[15] of the fourth kind, F_D, in several variables occurs as a result of the operation of $(L_{n-2}^{n-3})^{\alpha}$. These statements are obvious generalizations of the case $n = 4$. [See Eqs. (9)-(11) below.]

The fact that these Appell and Lauricella functions appear so rarely (if at all) elsewhere in theoretical physics perhaps warrants a systematic analysis of such functions, and generalizations thereof via the boson calculus of the $SU(n)$ state.

It is obvious that the combined action of products of lowering operators will lead to a multiple sum of products of folded blocks of terms. For example, at

the $SU(4)$ level, one already has to deal with a folded product of the Gauss, Appell, and Lauricella functions.

In analyzing the structure of these new functions beyond the known repertory, one criterion used is to check whether their integral representations admit a definite pattern of generalization. One class of generalized hypergeometric functions has been defined by Gel'fand *et al.*[16] as the Radon transforms[16] of products of *linear* forms. It so happens that all the known hypergeometric and generalized hypergeometric functions such as the Gauss, Appell, and Lauricella functions have this property.[17] The question arises whether this feature holds for all functions generated for the general $SU(n)$ states. This question was hitherto unsettled even for $n = 4$. We show that, in general, these combinatorially generated functions for the $SU(n)$ states are not confined to the class of generalized hypergeometric functions which are Radon transforms of products of linear forms. Already at the $SU(4)$ level, while special cases may possess such property, the general $SU(4)$ states do not. This in essence answers the question[5] posed by Ciftan in a negative way.

2. COMBINATORIAL STRUCTURE FOR GENERAL $SU(4)$ STATES

There are obvious advantages in reaching a general $SU(n)$ state by approximate application of lowering operators from the semimaximal state. Besides the apparent ease of getting the normalization constants,[4] a subject which will not concern us here, the structure of the semimaximal state is sufficiently simple so that the combinatorics ensued in pushing through the set of lowering operators can be systematically controlled.

We shall assume that the reader is reasonably familiar with the boson operator formalism[2] for the representations for the $U(n)$ groups. Thus we shall merely sketch the necessary expressions for the sake of getting the notations straight.

In terms of the creation and annihilation operators $a_i^{(\lambda)}$ and $\bar{a}_j^{(\lambda)}$, the elements of the Lie algebra read

$$E_{ij} = \sum a_i^{(\lambda)} \bar{a}_j^{(\lambda)}, \quad i, j = 1, \cdots, n. \tag{2}$$

The commutation relations

$$[E_{ij}, E_{rs}] = \delta_{jr} E_{is} - \delta_{is} E_{rj} \tag{3}$$

follow from the canonical commutation relations for the a's,

$$[\bar{a}_i^{(\lambda)}, a_j^{(\lambda')}] = \delta_{ij} \delta^{\lambda\lambda'}. \tag{4}$$

In terms of the (antisymmetrized) determinantal factors

$$a_{i_1 i_2 \cdots i_s} = \sum \epsilon(i_1 i_2 \cdots i_s) a_{i_1} a_{i_2} \cdots a_{i_s}, \tag{5}$$

where $\epsilon(i_1 i_2 \cdots i_s)$ is $+1$ if the set of indices is an even permutation of $(1, 2, \cdots, s)$ and is -1 otherwise, we have thus

$$\begin{aligned} &|\text{general } SU(4) \text{ state}\rangle \\ &\equiv \left| \begin{pmatrix} m_{14} & m_{24} & m_{34} & 0 \\ & m_{13} & m_{23} & m_{33} \\ & & m_{12} & m_{22} \\ & & & m_{11} \end{pmatrix} \right\rangle\!\!\rangle \\ &= \text{const}\,(L_2^1)^{n_{12}} (L_3^2)^{n_{23}} (L_3^1)^{n_{13}} \\ &\quad \times \left| \begin{pmatrix} m_{14} & m_{24} & m_{34} & 0 \\ & m_{13} & m_{23} & m_{33} \\ & & m_{13} & m_{23} \\ & & & m_{13} \end{pmatrix} \right\rangle\!\!\rangle \\ &= \text{const}\,(L_2^1)^{n_{12}} (L_3^2)^{n_{23}} (L_3^1)^{n_{13}} (a_{123})^{\nu_{33}} (a_{124})^{n_{34}} \\ &\quad \times (a_{12})^{\nu_{23}} (a_{14})^{n_{24}} (a_1)^{\nu_{13}} (a_4)^{n_{14}} |0\rangle, \end{aligned} \tag{6}$$

where the set of lowering operators are related to the E_{ij} as follows[3]:

$$L_k^{k-1} \equiv E_{k,k-1}, \quad k = 2, \cdots, n-1, \tag{7a}$$

$$L_k^{k-2} \equiv \mathcal{E}_{k-2,k-1} E_{k,k-2} + E_{k,k-1} E_{k-1,k-2}, \text{ etc.,} \tag{7b}$$

$$\mathcal{E}_{ij} \equiv (E_{ii} - E_{jj}) + (j - i)1. \tag{7c}$$

In the exponent, the following short-hand notation is used:

$$n_{ij} \equiv m_{ij} - m_{ij-1}, \quad \nu_{ij} \equiv m_{ij} - m_{i+1,j+1}. \tag{7d}$$

We note first the action of each operator separately. On the $SU(4)$ level, $(L_3^1)^{n_{13}}$ leads to the Appell function of the second kind F_2; $(L_3^2)^{n_{23}}$ gives the Gauss function ${}_2F_1$, and finally $(L_2^1)^{n_{12}}$ yields the Lauricella function of the fourth kind $F_D^{(3)}$ in 3-variables.[2,5] Since these are crucial ingredients, they have been rederived and recorded here below for the sake of readability. The defining power series expansions and their integral representations are given in the Appendix. We have

$$\begin{aligned} &(L_3^1)^{n_{13}} (a_{124})^{n_{34}} (a_{12})^{\nu_{23}} (a_{14})^{n_{24}} (a_1)^{\nu_{13}} |0\rangle \\ &= \frac{\Gamma(\nu_{13} + 1)\Gamma(s + 2)}{\Gamma(\nu_{13} - n_{13} + 1)\Gamma(s + 2 - n_{13})} \\ &\quad \times (a_{124})^{n_{34}} (a_{12})^{\nu_{23}} (a_{14})^{n_{24}} (a_1)^{\nu_{13} - n_{13}} (a_3)^{n_{13}} \\ &\quad \times F_2(-n_{13}; -n_{34}, -n_{24}; \\ &\qquad -s - 1, \nu_{13} - n_{13} + 1; w_1, w_2) |0\rangle, \end{aligned} \tag{8}$$

where $s = \nu_{13} + \nu_{23} + n_{24} + n_{34}$,

$$(L_3^2)^{n_{23}}(a_{124})^{n_{31}-k_1}(a_{12})^{\nu_{23}}|0\rangle = \frac{\Gamma(\nu_{23}+1)}{\Gamma(\nu_{23}-n_{23}+1)}(a_{124})^{n_{31}-k_1}(a_{12})^{\nu_{23}-n_{23}}(a_{13})^{n_{23}} \times {}_2F_1(-n_{23}, -n_{34}+k_1; \nu_{23}-n_{23}+1; w_3)|0\rangle, \quad (9)$$

$$(L_2^1)^{n_{12}}(a_{134})^{k_3}(a_{13})^{n_{23}-k_3}(a_{14})^{n_{24}-k_2}(a_1)^{\nu_{13}-n_{13}+k_2}|0\rangle = \frac{\Gamma(n_{23}+1)}{\Gamma(n_{23}-n_{12}+1)}\frac{(n_{12}-n_{23})_{k_3}}{(-n_{23})_{k_3}}(a_{134})^{k_3}(a_{13})^{n_{23}-n_{12}-k_3} \times (a_{14})^{n_{24}-k_2}(a_1)^{\nu_{13}-n_{13}+k_2}(a_{23})^{n_{12}} \times F_D^{(3)}(-n_{12}; -n_{24}+k_2, -\nu_{13}+n_{13}-k_2, -k_3; 1+n_{23}-n_{12}-k_3; w_4, w_5, w_6)|0\rangle, \quad (10)$$

where $(c)_r \equiv \Gamma(c+r)/\Gamma(c)$. The w's are defined in Eq. (13) below. Expressing the ${}_2F_1$, F_2, and $F_D^{(3)}$ in the standard single, double, and triple power-series expansions, respectively (see Appendix), we get from Eq. (6)

$$|\text{general } SU(4) \text{ state}\rangle = \text{const} \times (a_{123})^{\nu_{33}}(a_{124})^{n_{31}}(a_{12})^{\nu_{23}-n_{23}}(a_{13})^{n_{23}-n_{12}} \times (a_{14})^{n_{24}}(a_{23})^{n_{12}}(a_1)^{\nu_{13}-n_{13}}(a_3)^{n_{13}}(a_4)^{n_{14}}S^{(4)}|0\rangle, \quad (11)$$

$$S^{(4)} = \sum_{k_1,\cdots,k_6} \frac{(-n_{13})_{k_1+k_2}(-n_{34})_{k_1}(-n_{24})_{k_2}}{(-s-1)_{k_1}(1+\nu_{13}-n_{13})_{k_2}} \frac{(n_{12}-n_{23})_{k_3}(-n_{34}+k_1)_{k_3}}{(1+\nu_{23}-n_{23})_{k_3}} \times \frac{(-n_{12})_{k_4+k_5+k_6}(-n_{24}+k_2)_{k_4}(-\nu_{13}+n_{13}-k_2)_{k_5}(-k_3)_{k_6}}{(1+n_{23}-n_{12}-k_3)_{k_4+k_5+k_6}} \prod_{j=1}^{6}\frac{(w_j)^{k_j}}{k_j!}, \quad (12)$$

where

$$w_1 = \frac{a_{123}a_4}{a_{124}a_3}, \quad w_2 = \frac{a_1a_{34}}{a_3a_{14}}, \quad w_3 = \frac{a_{12}a_{134}}{a_{13}a_{124}}, \quad w_4 = \frac{a_{13}a_{24}}{a_{14}a_{23}}, \quad w_5 = \frac{a_2a_{13}}{a_1a_{23}}, \quad w_6 = \frac{a_{13}a_{234}}{a_{23}a_{134}}. \quad (13)$$

Equation (11) is essentially equivalent to Eq. (4.6c) of Ref. 5, apart from some obvious misprints there.

Using the standard integral representations[13,14] for those block constituents in (12), one can easily convert the sixfold summation into a fourfold integral representation[18]:

$$S^{(4)} = \text{const} \times \int_0^1\int\int\int dt_1\,dt_2\,dt_3\,dt_4 \times t_1^{-n_{34}-1}t_2^{-n_{24}-1}t_3^{-n_{23}-1}t_4^{-n_{12}-1} \times (1-t_1)^{n_{34}-s-2}(1-t_2)^{n_{24}+\nu_{13}-n_{13}}(1-t_3)^{\nu_{23}} \times (1-t_4)^{n_{23}}(1-w_4t_4)^{n_{24}}(1-w_5t_4)^{\nu_{13}-n_{13}} \times [1-w_3t_3(1-w_6t_4)/(1-t_4)]^{n_{34}} \times \{1-w_1t_1[1-w_3t_3(1-w_6t_4)/(1-t_4)]^{-1} - w_2t_2(1-w_5t_4)/(1-w_4t_4)\}^{n_{13}}. \quad (14)$$

By inspection, one easily convinces oneself that, because of the delicate folding of variables involved, it does not seem likely that Eq. (14) in general can be brought to be the Radon transform of products of *linear* forms. In the next section, we shall see that only special cases of Eq. (14) will have such a simple property.

3. RADON STRUCTURE OF THE SEMISEMIMAXIMAL $SU(4)$ STATE

This is another case which was started by Ciftan but was left unsettled. For the semisemimaximal state, i.e., the state with $n_{12}=0$, or $m_{12}=m_{11}$, the net effect is that the threefold summation involving the Lauricella function is now absent. The analog of Eq. (14) here is thus considerably simpler:

$$S^{(4)}(n_{12}=0, w_4=w_5=w_6=1) = \text{const} \times \int_0^1\int\int dt_1\,dt_2\,dt_3 t_1^{-n_{31}-1}t_2^{-n_{24}-1}t_3^{-n_{23}-1} \times (1-t_1)^{n_{34}-s-2}(1-t_2)^{n_{24}+\nu_{13}-n_{13}} \times (1-t_3)^{\nu_{23}}(1-w_3t_3)^{n_{34}} \times [1-w_1t_1/(1-w_3t_3)-w_2t_2]^{n_{13}}. \quad (15)$$

Equation (15) can be easily cast into the Radon form with the aid of the following change of variables:

$$x_1 = w_1t_1, \quad x_2 = 1-w_2t_2, \quad x_3 = (1-w_2t_2)(1-w_3t_3). \quad (16)$$

We have

$$S^{(4)}(n_{12}=0, w_4=w_5=w_6=1)$$
$$= \text{const} \times w_1^{s+2} w_2^{n_{13}-\nu_{13}} w_3^{n_{23}-\nu_{23}}$$
$$\times \iiiint dx_1\, dx_2\, dx_3\, dx_4 \delta\left(1-\sum_{i=1}^{4} x_i\right)$$
$$\times \prod_{k=1}^{9} (\xi^{(k)}, x)^{b_k}, \tag{17}$$

where

$$(\xi, x) \equiv \sum_{i=1}^{4} \xi_i x_i$$

denotes a linear form and the coefficients are

$$\xi^{(1)} = (1,0,0,0), \quad \xi^{(2)} = (0,1,0,0),$$
$$\xi^{(3)} = (0,0,1,0), \quad \xi^{(4)} = (-1,0,1,0),$$
$$\xi^{(5)} = (0,1,-1,0), \quad \xi^{(6)} = (0, w_3-1, 1, 0), \tag{18a}$$
$$\xi^{(7)} = (1,0,1,1), \quad \xi^{(8)} = (w_1-1, w_1, w_1, w_1),$$
$$\xi^{(9)} = (w_2-1, w_2, w_2-1, w_2-1),$$

$$b_1 = -n_{34}-1, \quad b_2 = n_{13}+n_{23}-n_{34}-\nu_{23},$$
$$b_3 = n_{34}-n_{13}, \quad b_4 = n_{13},$$
$$b_5 = -n_{23}-1, \quad b_6 = \nu_{23}, \tag{18b}$$
$$b_7 = -n_{24}-1, \quad b_8 = n_{34}-s-2,$$
$$b_9 = n_{24}+\nu_{13}-n_{13}.$$

The right-hand side of Eq. (17) is readily recognized as the Radon transform of products of linear forms in a 4-dimension space. However, as mentioned earlier, unfortunately the general case Eq. (14) does not share this property.

We thus conclude that the class of functions generated by the boson combinatorics, in general, is not confined to the class of generalized hypergeometric functions defined as Radon transforms of products of linear forms.

APPENDIX

For the sake of readability, we give here the relevant definitions and integral representations for the Gauss, Appell and Lauricella functions involved[13,14]:

(a) The Gauss function ${}_2F_1$:

$${}_2F_1(a, b; c; z)$$
$$= \sum_{k=0}^{\infty} \frac{(a)_k (b)_k}{(c)_k} \frac{z^k}{k!} \tag{A1}$$
$$= \frac{\Gamma(c)}{\Gamma(b)\Gamma(c-b)} \int_0^1 dt\, t^{b-1}(1-t)^{c-b-1}(1-zt)^{-a}; \tag{A2}$$

(b) The Appell function of the second kind, F_2:

$$F_2(a; b_1, b_2; c_1, c_2; z_1, z_2)$$
$$= \sum_{k_1,k_2} \frac{(a)_{k_1+k_2}(b_1)_{k_1}(b_2)_{k_2}}{(c_1)_{k_1}(c_2)_{k_2}} \frac{z_1^{k_1}}{k_1!} \frac{z_2^{k_2}}{k_2!} \tag{A3}$$
$$= \frac{\Gamma(c_1)\Gamma(c_2)}{\Gamma(b_1)\Gamma(c_1-b_1)\Gamma(b_2)\Gamma(c_2-b_2)}$$
$$\times \iint_0^1 dt_1\, dt_2 t_1^{b_1-1} t_2^{b_2-1}(1-t_1)^{c_1-b_1-1}$$
$$\times (1-t_2)^{c_2-b_2-1}(1-z_1t_1-z_2t_2)^{-a}; \tag{A4}$$

(c) The Lauricella function of the fourth kind, $F_D^{(n)}$, in n-variables:

$$F_D(a; b_1, \cdots, b_n; c; z_1, \cdots, z_n)$$
$$= \sum_{k_i} \frac{(a)_{\Sigma k_i} \prod_{j=1}^{n} (b_j)_{kj}}{(c)_{\Sigma k_i}} \prod_{j=1}^{n} \frac{z_j^{k_j}}{k_j!} \tag{A5}$$
$$= \frac{\Gamma(c)}{\Gamma(a)\Gamma(c-a)} \int_0^1 dt\, t^{a-1}(1-t)^{c-a-1} \prod_{i=1}^{n} (1-z_it)^{-b_i}. \tag{A6}$$

* Work supported in part by the U.S. Office of Naval Research.

[1] Among earlier references on the boson operator formalism, we mention the following: J. Schwinger, "On Angular Momentum," U.S. Atomic Energy Commission Report NYO-3071, 1952, reprinted in *Quantum Theory of Angular Momentum*, edited by L. C. Biedenharn and H. Van Dam (Academic, New York, 1965), p. 229; V. Bargmann and M. Moshinsky, Nucl. Phys. **18**, 697 (1960); **23**, 177 (1961); M. Moshinsky, *ibid.* **31**, 384 (1962); Rev. Mod. Phys. **34**, 813 (1962); J. Math. Phys. **4**, 1128 (1963); also Refs. 2–4 below.

[2] G. E. Baird and L. C. Biedenharn, J. Math. Phys. **4**, 1449 (1963).

[3] J. G. Nagel and M. Moshinsky, J. Math. Phys. **6**, 682 (1965).

[4] M. Ciftan and L. C. Biedenharn, J. Math. Phys. **10**, 221 (1969).

[5] M. Ciftan, J. Math. Phys. **10**, 1635 (1969).

[6] I.e., a $U(n)$ state such that it is a maximal state on the $U(n-1)$ level; a maximal state is one where the entries in the Gel'fand pattern take on maximal values $m_{i\,j-1} = m_{ij}$ for $j = 2, \cdots, n$ and $i = 1, \cdots, n-1$.

[7] I. M. Gel'fand and M. L. Zeltin, Doklady Akad. Nauk SSSR **71**, 825 (1950).

[8] For similar considerations in the extension to the groups $Sp(4)$ and $O(5)$, see, e.g., W. J. Holman III, J. Math. Phys. **10**, 1710 (1969).

[9] E. P. Wigner, "Applications of Group Theory to the Special Functions of Mathematical Physics," unpublished lecture notes, Princeton University, 1955.

[10] N. J. Vilenkin, *Special Functions and the Theory of Group Representations* (Transl. Math. Mono. Vol. 22) (American Mathematical Society, Providence, R.I., 1968).

[11] W. Miller, Jr., *Lie Theory and Special Functions* (Academic, New York, 1968).

[12] J. D. Talman, *Special Functions—A Group Theoretic Approach* (Benjamin, New York, 1968).

[13] P. Appell and J. Kampé de Fériet, *Fonctions hypergeometriques et hyperspheriques* (Gauthier-Villars, Paris, 1926), esp. Chap. 7; G. Lauricella, Rend. Circ. Mat. Palermo **7**, 111 (1893).

[14] L. J. Slater, *Generalized Hypergeometric Functions* (Cambridge U.P., Cambridge, 1966); esp. Chap. 8.

[15] The apparent twist in nomenclature may be unfortunate. The Lauricella function of the fourth kind, F_D, actually corresponds to the generalization of the Appell function of the first kind, F_1, to several variables.

[16] I. M. Gel'fand, M. I. Graev, and N. Ya. Vilenkin, *Generalized Functions* (Academic, New York, 1966), Vol. V.

[17] Except, of course, the Appell function of the fourth kind, F_4, for which no analogous integral representation is known.

[18] Here, the t_1, t_2 integrals come from the F_2 or the k_1, k_2 sum; the t_3 integral from the ${}_2F_1$ or the k_3 sum; finally the t_4 integral from $F_D^{(3)}$ or the k_4, k_5, k_6 sum.

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 12, NUMBER 3 MARCH 1971

Dual Transformations in Many-Component Ising Models*

L. MITTAG
Saybrook College, Yale University, New Haven, Connecticut
AND
M. J. STEPHEN†
Physics Department, Rutgers University, New Brunswick, New Jersey

(Received 17 August 1970)

Dual transformations in many-component Ising models in two dimensions on a square lattice are studied. The models considered include those of Ashkin and Teller and of Potts. In certain cases the dual transformation is a relation between the partition function of a lattice at high and low temperatures and can be used to determine a unique critical temperature if one exists. Dual transformations are considered both from a topological and an algebraic point of view. The topological arguments are a natural extension of those used by Onsager for the 2-component Ising model. The transfer matrices for these models are constructed, and it is shown how the dual transformation arises in this formulation of the problem. The algebras generated by these models are investigated and provide a generalization of the spinor algebra introduced by Kaufman in the 2-component Ising model.

1. INTRODUCTION

The Ising model is of great interest and importance in the theory of phase transitions. The original model, proposed by Ising,[1] is a 2-component model in which each site of a crystal lattice is occupied by one of two kinds of atoms. Only nearest-neighbor interactions are considered. Equivalently, each site could be regarded as occupied by a classical particle with spin $S = \frac{1}{2}$, each spin assuming two orientations "up" or "down" corresponding to the two kinds of atoms.

Dual transformations have played an important role in the development of the Ising model. Kramers and Wannier,[2] using a matrix approach to the 2-component Ising model on a square lattice in zero magnetic field, were able to derive a relation between the partition function (PF) at high and low temperatures. By assuming the existence of a single critical point, they were able to determine the critical temperature. Onsager,[3] using topological arguments, generalized the results of Kramers and Wannier to a wider class of 2-dimensional lattices. Onsager[3] and Kaufman,[4] using the transfer matrix approach, also indicated how the dual transformation arises in this formulation of the problem. In general, a dual transformation is a relation between the PF of one lattice with certain Boltzmann factors and the PF of another lattice with other Boltzmann factors. The transformation between the two sets of Boltzmann factors is reciprocal. In certain cases this transformation defines a relation between the PF of a lattice at high and low temperatures. This relation can then be used to determine a unique critical temperature if one exists.

A number of generalizations of the original Ising model have been considered. We will be concerned here with many-component models in which each site of a lattice is occupied by one of q different kinds of atoms. In 1943 Ashkin and Teller[5] considered a 4-component model on a square lattice on which each lattice site is occupied by one of four kinds of atoms. Using topological arguments, they demonstrated the existence of a dual transformation for their model and located the transition temperature in the special case where nearest-neighbor-like atoms have an interaction energy ϵ_0 and unlike atoms have an interaction energy ϵ_1 with $\epsilon_0 < \epsilon_1$. The low temperature state is thus an ordered one.

Several generalizations of the 2-component Ising model have been proposed by Potts[6] in which each lattice site is occupied by one of q different kinds of atoms. In one model (which we will call the Potts model) there are only two independent interaction energies. Nearest-neighbor-like atoms have an interaction energy ϵ_0 and nearest-neighbor-unlike atoms an interaction energy ϵ_1 with $\epsilon_0 < \epsilon_1$. Using the Kramers–Wannier matrix approach, he was able to demonstrate a relation between the PF at high and low temperatures and thus locate the transition temperature for all q. Fisher[7] has used a topological approach to obtain the dual transformation and hence to locate the transition temperature for the Potts model. Potts[6] also considered another model in which the q configurations of each lattice site are represented by a 2-dimensional vector which can point in any one of q symmetrically placed directions. We will refer to this model as the Potts vector model. Again using the matrix approach, he was able to locate the transition temperature for $q = 3$, 4. It was pointed out independently by Betts[8] and Suzuki[9] that the Potts vector model for the case $q = 4$ factorizes into two independent 2-component models and is thus exactly

soluble. In the case $q = 3$ the two Potts models are equivalent.

The purpose of this paper is to study the existence of dual transformations in many-component Ising models in two dimensions on a square lattice. The models considered include those of Ashkin and Teller and of Potts, but it is also indicated how their results can be generalized in certain directions. We have derived the dual transformations in these models both from a topological point of view and by construction of the transfer matrix. The topological arguments are a natural extension of those used by Onsager[3] for the 2-component Ising model and simplify those used by Fisher[7] in the case of the Potts model. The transfer matrix approach is also interesting, and the algebras generated by these models are investigated. These considerations are a generalization of spinor approach used by Kaufman[4] in the 2-component Ising model. In Sec. II we discuss the 3-component Potts model, first, from a topological point of view and, second, we construct the transfer matrix. It is interesting to note that this transfer matrix is a quadratic form in certain operators. The dual transformation is then a similarity transformation of these operators. In Sec. III we discuss the 4-component Ashkin–Teller model again from the topological viewpoint and also construct the transfer matrix. The methods used in these two examples are easily generalized to higher-component models, and these generalizations are discussed in Sec. IV.

2. THREE-COMPONENT SYSTEM—POTTS MODEL

A. High-Temperature Expansion

Consider a 2-dimensional square lattice the sites of which are occupied by one of three kinds of atoms, A, B, or C. We suppose that only nearest-neighbor atoms interact with energies ϵ_0 if the neighbors are identical and ϵ_1 if they are different with $\epsilon_0 < \epsilon_1$. At low temperatures the system will then be ordered. It is convenient to define Boltzmann factors α and β by

$$\alpha = e^{-\epsilon_0/kT}, \qquad \beta = e^{-\epsilon_1/kT}, \tag{2.1}$$

where T is the temperature and k is Boltzmann's constant. Three diagonal matrices are defined by

$$P^{(1)} = \begin{pmatrix} 1 & & \\ & 0 & \\ & & 0 \end{pmatrix}, \quad P^{(2)} = \begin{pmatrix} 0 & & \\ & 1 & \\ & & 0 \end{pmatrix}, \quad P^{(3)} = \begin{pmatrix} 0 & & \\ & 0 & \\ & & 1 \end{pmatrix}, \tag{2.2}$$

and we introduce projection operators $P_r^{(i)}$ with respect to the A, B, and C atoms at site r of the lattice. We may write the energy of interaction, E_{rs}, of atoms at adjacent sites r and s as an operator, using the obvious vector notation

$$\begin{aligned} E_{rs} &= \epsilon_1 - (\epsilon_1 - \epsilon_0)(P_r^{(1)}P_s^{(1)} + P_r^{(2)}P_s^{(2)} + P_r^{(3)}P_s^{(3)}) \\ &= \epsilon_1 - (\epsilon_1 - \epsilon_0)\mathbf{P}_r \cdot \mathbf{P}_s . \end{aligned} \tag{2.3}$$

The total energy of a given configuration of atoms over the lattice is given by $E_c = \sum_{(\mathrm{nn})} E_{rs}$, where the sum extends over the $2N$ nearest-neighbor (nn) pairs on a square lattice of N sites with periodic boundary conditions. The PF is given by

$$Z_N(\alpha, \beta) = \mathrm{Tr} \prod_{\mathrm{nn}} \exp\left(-\frac{E_{rs}}{kT}\right), \tag{2.4}$$

where the trace is taken over the 3^N-dimensional vector space of configurations. Since projection operators commute and have the property $(\mathbf{P}_r \cdot \mathbf{P}_s)^2 = (\mathbf{P}_r \cdot \mathbf{P}_s)$, a given factor in (2.4) may be written

$$\exp(-E_{rs}/kT) = \beta + (\alpha - \beta)\mathbf{P}_r \cdot \mathbf{P}_s . \tag{2.5}$$

This can be written in a more convenient form by introducing a diagonal matrix

$$\Omega = \begin{pmatrix} 1 & & \\ & \omega & \\ & & \omega^2 \end{pmatrix}, \quad \omega = e^{2\pi i/3} \tag{2.6}$$

and defining, as before, Ω_r with respect to the rth site. The projection operators may then be expressed by

$$\begin{aligned} 3P_r^{(1)} &= I + \Omega_r + \Omega_r^+, \\ 3P_r^{(2)} &= I + \omega^2\Omega_r + \omega\Omega_r^+, \\ 3P_r^{(3)} &= I + \omega\Omega_r + \omega^2\Omega_r^+, \end{aligned} \tag{2.7}$$

and

$$3\mathbf{P}_r \cdot \mathbf{P}_s = I + \Omega_r\Omega_s^+ + \Omega_r^+\Omega_s . \tag{2.8}$$

The PF may now be written

$$Z_N(\alpha, \beta) = \left(\frac{u^2}{3}\right)^N \mathrm{Tr} \prod_{\mathrm{nn}} \left(1 + \frac{x}{u}(\Omega_r\Omega_s^+ + \Omega_r^+\Omega_s)\right) \tag{2.9}$$

with

$$3^{\frac{1}{2}}u = \alpha + 2\beta, \qquad 3^{\frac{1}{2}}x = \alpha - \beta. \tag{2.10}$$

The only nonzero terms in the product (2.9) occur for terms like Ω_r^3, $\Omega_r\Omega_r^+$, and $(\Omega_r^+)^3$ since $\mathrm{Tr}\,\Omega_r^3 = 3$, etc., and the trace of all other types of terms is zero. We can represent each term in the expansion of the product (2.9) graphically as follows: Let $\Omega_r\Omega_s^+$ be represented by a directed bond from site s to site r, $s \cdot\!\!\longrightarrow\!\!\cdot\, r$. The above conditions imply that at each site we have as many arrows entering as leaving or that either three arrows enter or leave. Each arrow or bond contributes one factor of x/u. Of course, there may

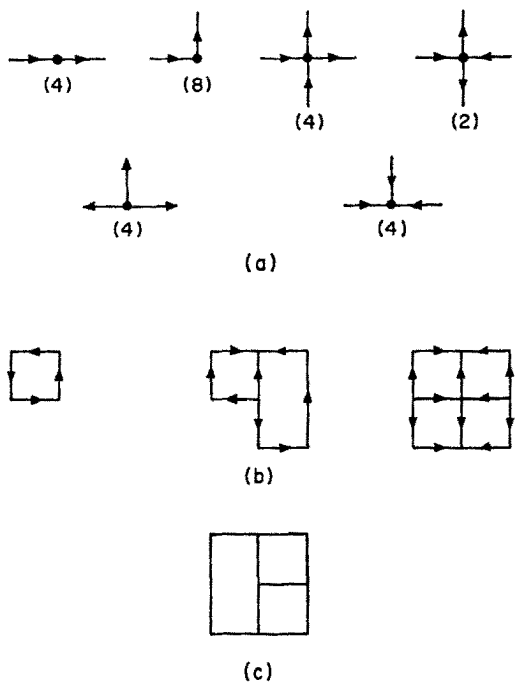

FIG. 1. (a) Possible arrangements of directed bonds at a lattice site. Numbers in parentheses indicate the number of possible configurations. (b) Examples of admissible diagrams. (c) An inadmissible diagram.

also be no bonds at a lattice site. The types of diagrams that can occur at each site are shown in Fig. 1(a). A diagram is considered admissible (i.e., contributes to the PF) if every lattice site is similar to one of those in Fig. 1(a). Examples of admissible diagrams are given in Fig. 1(b). These diagrams are members of a class called directed polygonal planar graphs.[10,11] A simple example of a planar graph which cannot be directed according to the above rules is shown in Fig. 1(c). Later we will show that all admissible diagrams can be face colored (including the exterior face) with no more than three colors such that no faces with an edge or bond in common have the same color. We will also show that every planar polygonal graph which can be three-colored can be directed so as to be admissible.

In the admissible diagrams we have a prescription for calculating an expansion of Z_N in powers of x/u. Thus we have obtained a functional representation $Z_N(\alpha, \beta) = u^{2N} f_N(x/u)$ with f_N defined by a power series. In the high temperature region $\alpha \sim \beta \sim 1$ and $x \ll u$. We assume that the power series defining f_N converges. The partition function per particle in the thermodynamic limit is given by

$$Z(\alpha, \beta) = \lim_{N \to \infty} [Z_N(\alpha, \beta)]^{1/N} = u^2 f(x/u), \tag{2.11}$$

where the second result applies at high temperatures.

B. Low-Temperature Expansion

If N_{AA} is the number of AA bonds, N_{AB} the number of AB bonds, etc., then the total energy of a given configuration L of the lattice is

$$E_L = \epsilon_0(N_{AA} + N_{BB} + N_{CC}) + \epsilon_1(N_{AB} + N_{AC} + N_{BC}). \tag{2.12}$$

The total number of bonds $2N$ satisfies

$$2N = N_{AA} + N_{BB} + N_{CC} + N_{AB} + N_{AC} + N_{BC}.$$

The Boltzmann factor corresponding to E_L is

$$\exp(-E_L/kT) = \alpha^{2N}(\beta/\alpha)^{N_{AB}+N_{AC}+N_{BC}}, \tag{2.13}$$

and the PF can be written as a power series in β/α,

$$Z_N(\alpha, \beta) = 3\alpha^{2N} \sum_l g_N(l)(\beta/\alpha)^l, \tag{2.14}$$

where $g_N(l)$ is the number of configurations with $l = N_{AB} + N_{BC} + N_{AC}$. The factor of 3 arises because the lowest-energy state $E = 2N\epsilon_0$ is triply degenerate. If we start from a configuration of all A atoms, then, as we change A atoms to B or C atoms, we generate a power series in β/α. Clearly this is a low-temperature expansion for $\epsilon_0 < \epsilon_1$, $\lim \beta/\alpha \to 0$ as $T \to 0$. Then from (2.11) and (2.14), in the thermodynamic limit, the partition function has the following functional representation:

$$\begin{aligned} Z(\alpha, \beta) &= \alpha^2 g(\beta/\alpha), \qquad \beta \ll \alpha, \\ &= u^2 f(x/u), \qquad \beta \sim \alpha. \end{aligned} \tag{2.15}$$

The existence of the dual transformation is related to the equivalence $f(x) = g(x)$. If this equality holds, then one can imagine constructing a lattice with Boltzmann factors u and x related to α and β by the reciprocal relations (2.10). The new lattice will have a PF $Z(u, x)$, $0 < x < u$, and the PF's of the two lattices will be equal:

$$Z(u, x) = Z(\alpha, \beta).$$

A topological proof, similar in spirit to Onsager's original proof with respect to the 2-component Ising model, yields the necessary equality of the PF's. Imagine an infinite face-centered square lattice in two dimensions. This lattice separates into two interpenetrating square lattices I and II. Label the lattice sites of I with A, B, and C according to some configuration L. The sites of lattice II will be unlabeled. For every AB, BC, or AC bond in I, draw a perpendicular bond in II. There is now drawn in II a planar polygonal graph. This graph can be directed to become admissible as follows. Choose an order for the atoms say ABCAB. Suppose one is standing on a site of I and

looking across a bond of II at a neighboring site. Then, if the atom at which one is looking follows the atom on which one is standing in the chosen order, direct the bond to the left. If it reverses the order, direct the bond to the right [see Fig. 2(a)]; if the two atoms are the same, of course, there is no bond. This procedure has been used recently by Baxter[12] in connection with the problem of coloring a hexagonal lattice. Some examples of situations that can occur are given in Fig. 2(b). All possible configurations on I now yield admissible diagrams on II. Note that if in an admissible diagram all arrows are reversed, another admissible diagram results. The transformations $A \to B \to C \to A$ and $A \to C \to B \to A$ and the identity all yield the same admissible diagram while the three transpositions $A \leftrightarrow B$, $B \leftrightarrow C$, and $C \leftrightarrow A$ yield the diagram with all bonds reversed. There is thus an admissible diagram for every three configurations. If we consider only finite diagrams (graphs with a finite number of bonds) and choose the infinite face to contain only A sites, there is a one-to-one relationship between configurations and admissible diagrams. Actually, we have only shown that for every configuration there is an admissible diagram. We must show the reverse, that for every admissible diagram there is a configuration on I. Equivalently, we can show that every admissible diagram can be face colored with at most three colors. This is proved in Appendix A.

In the limit of large N, the coefficients of x/u and β/α in the two expansions of $Z_N(\alpha, \beta)$, (2.9) and (2.14), become equal. We obtain the result

$$\begin{aligned} Z(\alpha, \beta) &= \alpha^2 f(\beta/\alpha), \quad \beta \ll \alpha, \\ &= u^2 f(x/u), \quad \beta \sim \alpha. \end{aligned} \tag{2.16}$$

The result is that, for every lattice constructed with Boltzmann factors α and β, there is a dual lattice constructed from Boltzmann factors u and x, the transformation being reciprocal. Alternatively, we may regard (2.16) as giving a relation between the high- and low-temperature behavior of the PF. If there is a unique critical temperature, it must occur when $x/u = \beta/\alpha$ with the solution $\alpha/\beta = 1 + \sqrt{3}$, a result first given by Potts.[6] For the 2-component Ising model $\alpha/\beta = 1 + \sqrt{2}$.

The 2-component Potts model is related to the problem of the residual entropy of 2-dimensional square ice. Topologically, the square ice problem can be reduced to counting the number of configurations W on a square lattice in which all nearest-neighbor vertices are joined by directed bonds. The ice condition at each vertex is that exactly two arrows point into each vertex and two arrows point away from each vertex. This set of configurations is a subset of the diagrams generated by (2.9). The number of configurations in the ice problem can be written

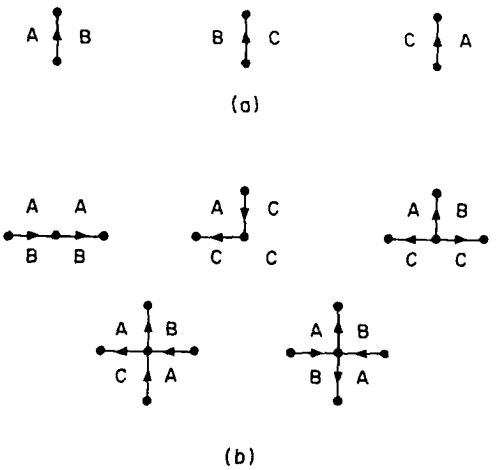

FIG. 2. (a) Rules for directing bonds separating different components. (b) Examples of possible arrangements of components at a vertex.

$$W = 3^{-N} \operatorname{Tr} \prod_{nn} (\Omega_r \Omega_s^+ + \Omega_r^+ \Omega_s). \tag{2.17}$$

The residual entropy has been evaluated exactly by Lieb.[13]

C. The Transfer Matrix

The 2-particle transfer matrix is given by

$$\begin{aligned} V_{ij} = \exp\left(-\frac{E_{ij}}{kT}\right) &= \begin{bmatrix} \alpha & \beta & \beta \\ \beta & \alpha & \beta \\ \beta & \beta & \alpha \end{bmatrix} \\ &= (\alpha - \beta) I + 3\beta T. \end{aligned} \tag{2.18}$$

The indices i and j represent the states A, B, and C. The last relation defines a 3×3 matrix $3T$, all of whose entries are unity. The eigenvalues of V_{ij} are $\alpha + 2\beta$ and $\alpha - \beta$, the last being degenerate. Thus the partition function of an $n \times 1$ lattice is $Z_{n,1} = (\alpha + 2\beta)^n + 2(\alpha - \beta)^n$.

The fact that T is a projection operator enables us to write V_{ij} in exponential form

$$V_{ij} = (\alpha - \beta) \exp(3H^* T), \tag{2.19}$$

with

$$e^{3H^*} = (\alpha + 2\beta)/(\alpha - \beta) = u/x. \tag{2.20}$$

Now imagine that an $n \times (m - 1)$ lattice is extended by adding a new column of atoms. The transfer matrix representing the interaction along the rows

between the last two columns is

$$V_1'(H^*) = (\alpha - \beta)^n \exp\left(3H^* \sum_{r=1}^{n} T_r\right), \quad (2.21)$$

where the index r refers to the row number. We now turn on the interaction of the atoms in the last column. The transfer matrix corresponding to this interaction is

$$V_2'(H) = \beta^n \exp\left(3H \sum_{r=1}^{n} \mathbf{P}_r \cdot \mathbf{P}_{r+1}\right), \quad (2.22)$$

where $3HkT = \epsilon_1 - \epsilon_0$. The partition function of an $n \times m$ lattice is now

$$\begin{aligned} Z_{nm}(\alpha, \beta) &= \operatorname{Tr}(V_2'V_1')^m \\ &= [\beta(\alpha - \beta)]^{nm} \operatorname{Tr}(e^{HA'}e^{H^*B'})^m, \end{aligned} \quad (2.23)$$

where

$$A' = 3\sum_{r=1}^{n} \mathbf{P}_r \cdot \mathbf{P}_{r+1}, \quad B' = 3\sum_{r=1}^{n} T_r. \quad (2.24)$$

Both T_r and $\mathbf{P}_r \cdot \mathbf{P}_{r+1}$ are projection operators of the same rank. One suspects that there is an algebraic transformation, in analogy with the 2-component Ising model, such that $T_r \to \mathbf{P}_r \cdot \mathbf{P}_{r+1} \to T_{r+1} \to \cdots$. If this is true, then the maximum eigenvalue of $\exp(HA')\exp(H^*B')$ is symmetric in H and H^*. If again a unique critical point exists, it would occur when $H = H^*$, which is equivalent to $\beta/\alpha = (\alpha - \beta)/(\alpha + 2\beta)$. We shall show that there exists a subspace in which all the eigenvalues of the above operator are symmetric in H and H^*.

Consider the matrix

$$M = \begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 1 & 0 & 0 \end{pmatrix}, \quad (2.25)$$

and define M_r in the usual direct product fashion

$$M_r = I \times I \times \cdots M \times I \times \cdots, \quad (2.26)$$

where M occupies the rth position and the I are unit matrices of order 3×3. We can then write

$$3T_r = 1 + M_r + M_r^+ \quad (2.27)$$

and, in place of (2.24), introduce the operators

$$A = \sum_r \Omega_r\Omega_{r+1}^+ + \Omega_r^+\Omega_{r+1}, \quad B = \sum_r M_r + M_r^+ \quad (2.28)$$

and

$$V_2(H) = \exp(HA), \quad V_1(H^*) = \exp(H^*B). \quad (2.29)$$

Apart from a multiplicative constant, the transfer matrix is now V_2V_1.

The matrices M and Ω satisfy the relations

$$M\Omega = \omega\Omega M, \quad M\Omega^+ = \omega^2\Omega^+M \quad (2.30)$$

and two others obtained by taking Hermitian conjugates. The group of order 27 generated by M and Ω under multiplication consists of all matrices of the form $\omega^l M^m \Omega^n$, $0 \le l, m, n \le 2$. If we consider instead the algebra generated by the products, there are 18 linearly independent elements over the reals or nine linearly independent elements over the complexes. We have then a complete matrix algebra.

We now define the following $3^n \times 3^n$ representation:

$$\begin{aligned} \pi_1 &= \Omega \times I \times I \times \cdots, \\ \pi_2 &= M^+ \times \Omega \times I \times \cdots, \\ \pi_3 &= M^+ \times M^+ \times \Omega \times \cdots, \\ &\cdots; \end{aligned} \quad (2.31)$$

$$\begin{aligned} Q_1 &= \Omega^+M \times I \times I \times \cdots, \\ Q_2 &= M \times \Omega^+M \times I \times \cdots, \\ Q_3 &= M \times M \times \Omega^+M \times \cdots, \\ &\cdots; \\ U &= M \times M \times M \times \cdots. \end{aligned} \quad (2.32)$$

Then

$$\begin{aligned} \pi_rQ_r &= M_r, \\ Q_r\pi_{r+1} &= \Omega_r^+\Omega_{r+1}, \quad 1 \le r \le n, \\ \pi_1Q_nU^2 &= \Omega_1\Omega_n^+. \end{aligned} \quad (2.33)$$

The algebra of the π, Q matrices is discussed in Appendix B.

The matrices A and B can now be written

$$\begin{aligned} A &= \sum_{r=1}^{n-1}(Q_r\pi_{r+1} + \pi_{r+1}^+Q_r^+) + \pi_1Q_nU^2 + Q_n^+\pi_1^+U, \\ B &= \sum_{r=1}^{n} \pi_rQ_r + Q_r^+\pi_r^+. \end{aligned} \quad (2.34)$$

Now introduce three projection operators

$$\begin{aligned} 3\Lambda_1 &= I + U + U^2, \\ 3\Lambda_2 &= I + \omega^2U + \omega U^2, \\ 3\Lambda_3 &= I + \omega U + \omega^2U^2, \end{aligned} \quad (2.35)$$

which have the properties

$$\begin{aligned} \Lambda_i\Lambda_j &= \delta_{ij}\Lambda_j, \\ \Lambda_1 + \Lambda_2 + \Lambda_3 &= I. \end{aligned} \quad (2.36)$$

Further, if $F(U)$ is a function of U and operators that commute with U, then

$$\begin{aligned} \Lambda_1F(U) &= F(I), \\ \Lambda_2F(U) &= F(\omega I), \\ \Lambda_3F(U) &= F(\omega^2 I). \end{aligned} \quad (2.37)$$

Note that A and B in (2.34) contain only matrices which commute with U, and thus we can separate the transfer matrix V_2V_1 into three parts

$$V_2V_1 = \Lambda_1V_2V_1 + \Lambda_2V_2V_1 + \Lambda_3V_2V_1, \quad (2.38)$$

which operate in different subspaces. In the awkward last factor in A, the U matrix is replaced by 1, ω, and ω^2, respectively. Consider $\Lambda_1V_2V_1$. The dual transformation D may be represented as[14]

$$D:\pi_1 \to Q_1 \to \pi_2 \to Q_2 \cdots Q_n \to \omega^2\pi_1 U. \quad (2.39)$$

Under this transformation,

$$\Lambda_1 \to \Lambda_1, \quad \Lambda_2 \leftrightarrow \Lambda_3,$$
$$\Lambda_1 A \to B, \qquad B \to \Lambda_1 A. \quad (2.40)$$

But D is an automorphism[15] of a complete matrix algebra and is thus equivalent to a similarity transformation which preserves eigenvalues, traces, etc.

For low temperatures the transfer matrix is dominated by $V_2(H) = e^{HA}$. The operator A is triply degenerate but is completely reduced by the projection operators Λ_i. Eigenvectors corresponding to degenerate eigenvalues appear in different subspaces. For high temperatures the result is different. Here the dominant factor is $V_1(H^*) = \exp H^*B$. The operator B, while similar to Λ_1A, has its maximum eigenvalue in the subspace spanned by Λ_1. Thus for low temperatures we expect to find three asymptotically degenerate maximum eigenvalues while at high temperatures a single maximum eigenvalue. It is well known that the degeneracy of the maximum eigenvalue of the transfer matrix is a necessary condition for long-range order.

3. FOUR COMPONENT SYSTEM, ASHKIN-TELLER MODEL

We again consider a 2-dimensional square lattice the sites of which are occupied by one of four kinds of atoms A, B, C, or D. In the Ashkin–Teller model only nearest-neighbor interactions are considered and the pair energies are specified as follows:

AA BB	AB	AC	AD	
CC DD	CD	BD	BC	(3.1)
ϵ_0	ϵ_1	ϵ_2	ϵ_3 .	

We also introduce the parameters

$$\alpha = e^{-\epsilon_0/kT}, \quad \beta = e^{-\epsilon_1/kT},$$
$$\gamma = e^{-\epsilon_2/kT}, \quad \delta = e^{-\epsilon_3/kT}. \quad (3.2)$$

The two Potts models for $q = 4$ are special cases of this model corresponding to (i) $\epsilon_1 = \epsilon_2 = \epsilon_3$ and (ii) $\epsilon_0 = -\epsilon_2$, $\epsilon_1 = \epsilon_3 = 0$, respectively.

Ashkin and Teller[5] demonstrated that a dual transformation exists for this model, and located the critical temperature. They used a topological argument, which generalized a method first used by Onsager. Betts[8] also studied this model from a topological point of view and pointed out that it is exactly soluble in the case where $\epsilon_0 + \epsilon_2 = \epsilon_1 + \epsilon_3$ which includes the Potts vector model. In this case the model factors into two independent 2-component Ising models. This was also noted by Suzuki.[9] We will briefly review the topological arguments and also construct the transfer matrix for this model.

The analysis of the 3-component model given in the previous section is readily extended to this 4-component model. However, it is convenient to proceed somewhat differently in a way which brings out more clearly the relation of this model with the 2-component Ising model and which is easily generalized to 2^M-component models. We introduce the following matrices:

$$S = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}, \qquad C = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \quad (3.3)$$

and

$$S^{(1)} = I \times S, \quad S^{(2)} = S \times I, \quad S^{(3)} = S \times S,$$
$$M^{(1)} = I \times C, \quad M^{(2)} = C \times I, \quad M^{(3)} = C \times C, \quad (3.4)$$

where I is a 2×2 unit matrix. We now introduce the usual direct-product representation $S_r^{(i)}$ and $M_r^{(i)}$, where r refers to a site and $i = 1, 2, 3$. A factor of the partition function is now

$$\exp(-E_{rs}/kT) = \tfrac{1}{2}(u + xS_r^{(1)}S_s^{(1)} + yS_r^{(2)}S_s^{(2)} + zS_r^{(3)}S_s^{(3)}) \quad (3.5)$$

with

$$2\alpha = u + x + y + z, \quad 2\beta = u - x + y - z,$$
$$2\gamma = u + x - y - z, \quad 2\delta = u - x - y + z. \quad (3.6)$$

The matrix of this reciprocal relation is $(C + S) \times (C + S)$, which is easily generalized to 2^M-component models. The PF is now

$$Z_N(\alpha, \beta, \gamma, \delta) = 4^{-N}\,\mathrm{Tr}\prod_{nn}(u + xS_r^{(1)}S_s^{(1)} + yS_r^{(2)}S_s^{(2)} + zS_r^{(3)}S_s^{(3)})$$
$$= u^{2N}F(x/u, y/u, z/u). \quad (3.7)$$

The function F is a symmetrical function of its arguments. If $uz = xy$, the above expression factors into two 2-component Ising models.

Equation (3.7) can be analyzed diagrammatically as noted by Ashkin and Teller. A set of planar polygonal graphs is obtained whose bonds fall into three classes $\{x, y, z\}$. At any vertex excluding an isolated vertex the following situations can occur: x^4, y^4, z^4, x^2y^2, y^2z^2, z^2x^2, x^2, y^2, z^2, and xyz. Polygonal graphs which can be so labeled are referred to by Ashkin and Teller as effective patterns.

Having assumed that $\alpha > \beta, \gamma, \delta$, Ashkin and Teller obtained a low-temperature expansion. The arguments are similar to those outlined in the previous section. It is found that the PF is given by the following expansion:

$$Z_N(\alpha, \beta, \gamma, \delta) = 4\alpha^{2N} \sum_{lmn} g_N(l, m, n) \left(\frac{\beta}{\alpha}\right)^l \left(\frac{\gamma}{\alpha}\right)^m \left(\frac{\delta}{\alpha}\right)^n = \alpha^2 G\left(\frac{\beta}{\alpha}, \frac{\gamma}{\alpha}, \frac{\delta}{\alpha}\right), \tag{3.8}$$

where $g_N(l, m, n)$ is the number of configurations with l AB pairs, m AC pairs, and n AD pairs. G is again a symmetric function of its arguments. The dual transformation consists in showing that $F \equiv G$. Then the PF of lattices constructed with Boltzmann factors $\alpha, \beta, \gamma, \delta$ and u, x, y, z are equal. This has been discussed in some detail by Ashkin and Teller. In the case where $\beta = \gamma = \delta$ from (3.6), we obtain the reciprocal relation

$$\beta/\alpha = (\alpha - \beta)/(\alpha + 3\beta), \tag{3.9}$$

which leads to a critical temperature determined by $\alpha/\beta = 1 + \sqrt{4}$.

The Transfer Matrix

Again we start with the 2-particle transfer matrix

$$V_{ij} = \begin{pmatrix} \alpha & \beta & \gamma & \delta \\ \beta & \alpha & \delta & \gamma \\ \gamma & \delta & \alpha & \beta \\ \delta & \gamma & \beta & \alpha \end{pmatrix} = \alpha + \beta M^{(1)} + \gamma M^{(2)} + \delta M^{(3)}. \tag{3.10}$$

This matrix can also be written $2(u\tau^{(0)} + x\tau^{(1)} + y\tau^{(2)} + z\tau^{(3)})$, which defines the matrices $\tau^{(i)}$, which turn out to be projection operators of the form $2^{-2}(I \pm C) \times (I \pm C)$. If we use these operators, it becomes simple to write the transfer matrix as

$$V_{ij} = A \exp(K_1^* M^{(1)} + K_2^* M^{(2)} + K_3^* M^{(3)}), \tag{3.11}$$

where

$$A^4 = 2^4 uxyz, \quad e^{4K_1^*} = uy/xz, \quad e^{4K_2^*} = ux/yz, \quad e^{4K_3^*} = uz/xy. \tag{3.12}$$

Aside from a factor A^n the transfer matrix representing the interaction along rows between successive columns is (using an obvious vector notation)

$$V_1(\mathbf{K}^*) = \exp\left(\sum_{i=1}^{3} \sum_r K_i^* M_r^{(i)}\right) = \exp\left(\sum_r \mathbf{K}^* \cdot \mathbf{M}_r\right). \tag{3.13}$$

Within a column, the transfer matrix, aside from a factor $(\alpha\beta\gamma\delta)^{n/4}$, is

$$V_2(\mathbf{K}) = \exp\left(\sum_{i=1}^{3} \sum_r K_i S_r^{(i)} S_{r+1}^{(i)}\right), \tag{3.14}$$

where

$$e^{4K_1} = \alpha\gamma/\beta\delta, \quad e^{4K_2} = \alpha\beta/\gamma\delta, \quad e^{4K_3} = \alpha\delta/\beta\gamma. \tag{3.15}$$

The PF is then given by

$$Z_{mn} = 2^{mn}(\alpha\beta\gamma\delta uxyz)^{mn/4} \operatorname{Tr}[V_2(\mathbf{K}) V_1(\mathbf{K}^*)]^m. \tag{3.16}$$

Since the relation between $\mathbf{K}$ and $\mathbf{K}^*$ is also reciprocal, the existence of a dual transformation is equivalent to the maximum eigenvalue of V_2V_1 being a symmetric function of $\mathbf{K}$ and $\mathbf{K}^*$, i.e., $\lambda_{\max}(\mathbf{K}, \mathbf{K}^*) = \lambda_{\max}(\mathbf{K}^*, \mathbf{K})$.

We introduce the following $4^n \times 4^n$ representation which is a generalization of the spinor representation introduced by Kaufman[4]:

$$\begin{aligned} P_1 &= S^{(1)} \times I \times I \cdots, \\ P_2 &= M^{(1)} \times S^{(1)} \times I \cdots, \\ &\cdots; \\ Q_1 &= -iS^{(1)}M^{(1)} \times I \times I \cdots, \\ Q_2 &= -iM^{(1)} \times S^{(1)}M^{(1)} \times I \cdots, \\ &\cdots; \\ R_1 &= S^{(2)} \times I \times I \cdots, \\ R_2 &= M^{(2)} \times S^{(2)} \times I \cdots, \\ &\cdots; \\ T_1 &= -iS^{(2)}M^{(2)} \times I \times I \cdots, \\ T_2 &= -iM^{(2)} \times S^{(2)}M^{(2)} \times I \cdots, \\ &\cdots. \end{aligned} \tag{3.17}$$

Here the I are 4×4 unit matrices. The P_i, Q_i and the R_i, T_i form spinor algebras in different spaces and so commute with each other. Using these operators, we can easily show that

$$M_r^{(1)} = iP_rQ_r, \quad M_r^{(2)} = iR_rT_r, \quad M_r^{(3)} = M_r^{(1)}M_r^{(2)}, \tag{3.18a}$$

$$S_r^{(1)}S_{r+1}^{(1)} = iQ_rP_{r+1}, \quad S_r^{(2)}S_{r+1}^{(2)} = iT_rR_{r+1}, \quad S_r^{(3)}S_{r+1}^{(3)} = S_r^{(1)}S_{r+1}^{(1)}S_r^{(2)}S_{r+1}^{(2)}, \tag{3.18b}$$

$$S_n^{(1)}S_1^{(1)} = iP_1Q_nU_1, \quad S_n^{(2)}S_1^{(2)} = iR_1T_nU_2,$$

where

$$U^{(1)} = M^{(1)} \times M^{(1)} \times \cdots,$$
$$U^{(2)} = M^{(2)} \times M^{(2)} \times \cdots.$$

The transfer matrices (3.13) and (3.14) can then be expressed in terms of the spinors (3.17) and take a form analogous to the 2-component Ising model. If we write $V_2 = e^A$ and $V_1 = e^B$, then

$$A = i\sum_{r=1}^{n-1}(K_1Q_rP_{r+1} + K_2T_rR_{r+1} + iK_3Q_rP_{r+1}T_rR_{r+1}) + iK_1P_1Q_nU_1 + iK_2R_1T_nU_2 - K_3P_1Q_nR_1T_nU_1U_2, \tag{3.19}$$

$$B = i\sum_{r=1}^{n}(K_1^*P_rQ_r + K_2^*R_rT_r + iK_3^*P_rQ_rR_rT_r). \tag{3.20}$$

We now introduce projection operators

$$\Lambda_{\pm\pm} = \tfrac{1}{4}(I \pm U_1)(I \pm U_2) \tag{3.21}$$

and consider $\Lambda_{++}V_2V_1 = V_2^{++}V_1$. These projection operators split the space into even and odd parts and eliminate the awkward end factors in (3.19). The dual transformation can be represented by

$$D\begin{cases} P_1 \to Q_1 \to P_2 \cdots Q_n \to -P_1 \\ R_1 \to T_1 \to R_2 \cdots T_n \to -R_1 \end{cases}. \tag{3.22}$$

Under this transformation, we have

$$\Lambda_{++}V_2(\mathbf{K})V_1(\mathbf{K}^*) \to \Lambda_{++}V_1(\mathbf{K})V_2(\mathbf{K}^*).$$

Since the algebra generated by P, Q, R, and T is a complete matrix algebra, the automorphism D can be represented by a similarity transformation[15] preserving eigenvalues, etc. Thus the eigenvalues of $V_2^{++}V_1$ are symmetric in $\mathbf{K}$ and $\mathbf{K}^*$. They are also real since $V_1^{\frac{1}{2}}V_2V_1^{\frac{1}{2}}$ is Hermitian and positive definite.

The matrices A and B contain quartic terms unless $K_3 = K_3^* = 0$. These conditions imply the equations

$$\alpha\delta = \beta\gamma, \quad uz = xy, \quad \epsilon_0 + \epsilon_3 = \epsilon_1 + \epsilon_2. \tag{3.23}$$

The transfer matrix then factors into two commuting operators which can be diagonalized separately. [The model is symmetrical in ϵ_1, ϵ_2, and ϵ_3, and the same factorization results when the relations (3.23) are permuted.]

In the special case $\epsilon_0 < \epsilon_1 = \epsilon_2 = \epsilon_3$, the matrix

$$A = \sum_r \sum_{i=1}^{3} S_r^{(i)}S_{r+1}^{(i)}$$

is quadruply degenerate, and at low temperatures one expects the largest eigenvalue to be quadruply degenerate in the thermodynamic limit. The projection operators $\Lambda_{\pm\pm}$ again split this degeneracy. At high temperatures the largest eigenvalue is expected to be nondegenerate and to occur in the space of Λ_{++}.

4. GENERALIZATIONS

A. The q-Component Potts Model

The projection operators $P_r^{(j)}$ readily generalize to the case of q components. Similarly the matrices M and Ω become, with $\omega = \exp(2\pi i/q)$,

$$\Omega = \begin{pmatrix} 1 & & & & & \\ & \omega & & & & \\ & & \omega^2 & & & \\ & & & \cdot & & \\ & & & & \cdot & \\ & & & & & \omega^{q-1} \end{pmatrix},$$

$$M = \begin{pmatrix} 0 & 1 & 0 & \cdots \\ 0 & 0 & 1 & \cdots \\ \cdot & \cdot & \cdot & \\ \cdot & \cdot & \cdot & \cdots \\ \cdot & \cdot & \cdot & \\ 1 & 0 & 0 & \cdots \end{pmatrix}. \tag{4.1}$$

Clearly

$$qP_r^{(j)} = \sum_{k=1}^{q} \omega^{-(j-1)k}\Omega_r^k. \tag{4.2}$$

If we use the methods of Sec. 2, it is easy to show that the transfer matrix can again be written as the product of $V_2(H) = \exp HA$ and $V_1(H^*) = \exp H^*B$, with

$$A = \sum_r \sum_{k=1}^{q-1} \Omega_r^k\Omega_{r+1}^{q-k}, \quad qHKT = \epsilon_1 - \epsilon_0,$$
$$B = \sum_r \sum_{k=1}^{q-1} M_r^k, \quad \exp qH^* = [\alpha + (q-1)\beta]/(\alpha - \beta). \tag{4.3}$$

The reciprocal relations are replaced by

$$q^{\frac{1}{2}}u = \alpha + (q-1)\beta, \quad q^{\frac{1}{2}}x = \alpha - \beta, \tag{4.4}$$

and the transition temperature is determined by $H = H^*$ or $\beta/\alpha = u/x$. This equation has the solution

$$\exp[(\epsilon_1 - \epsilon_0)/kT_c] = 1 + q^{\frac{1}{2}} \tag{4.5}$$

which was first given by Potts.[6] It is interesting to note that as $q \to \infty$, then $T_c \to 0$. The energy may also be evaluated at T_c, and this is carried out in Appendix C.

The algebraic representation defined by Eqs. (2.31) and (2.32) may be used as written by interpreting M and Ω as the $q \times q$ matrices in (4.1). If we consider only the subspace spanned by the projection operator

$$\Lambda_1 = q^{-1}\sum_{k=1}^{q} U^k,$$

then the operator $\Lambda_1V_2(H)V_1(H^*)$ is symmetric in H

and H^*. The dual transformation in this case is given by

$$D:\pi_1 \to Q_1 \to \pi_2 \cdots \to Q_n \to \omega^{-1}\pi_1 U^{-2}, \quad (4.6a)$$

$$D:U \to U^{-1}, \quad D:\Lambda_1 \to A_1. \quad (4.6b)$$

The last factor in (4.6a) is just sufficient to preserve the commutation relations $[\pi_n, Q_n]$ and $[Q_n, \pi_1]$ under the transformation. Thus D is an automorphism and $\Lambda_1 V_2(H)V_1(H^*)$ is similar to $\Lambda_1 V_1(H)V_2(H^*)$.

B. The Potts Vector Model

The states here are represented by q 2-dimensional vectors $\mathbf{e}^{(l)}$ which make angles $\theta_l = 2\pi l/q$, $l = 0, 1, \cdots, q-1$, with the positive x axis. The energy of nearest neighbors in states l and m is given by

$$E_{rs} = -J\mathbf{e}_r^{(l)} \cdot \mathbf{e}_s^{(m)}. \quad (4.7)$$

This can be written as an operator

$$E_{rs} = -J \sum_{l,m} P_r^{(l)} P_s^{(m)} \cos\theta_{l-m}, \quad (4.8)$$

which simplifies to

$$E_{rs} = -\tfrac{1}{2}J(\Omega_r\Omega_s^+ + \Omega_r^+\Omega_s). \quad (4.9)$$

The PF takes the simple form, with $2HKT = J$,

$$Z_N = \mathrm{Tr} \prod_{\mathrm{nn}} \exp\left[H(\Omega_r\Omega_s^+ + \Omega_r^+\Omega_s)\right]. \quad (4.10)$$

Similarly the factor V_2 of the transfer matrix is

$$V_2 = \exp\left(H \sum_r (\Omega_r\Omega_{r+1}^+ + \Omega_r^+\Omega_{r+1})\right), \quad (4.11)$$

while the other factor may be shown to be

$$V_1 = \exp\left(\sum_r \sum_{j=1}^{q-1} \gamma_j M_r^j\right). \quad (4.12)$$

There is now no simple symmetry relation between V_2 and V_1 except for $q = 2, 3, 4$.

C. The 2^M Ashkin–Teller Model

The Ashkin–Teller model can be generalized in the case where the number of components at each site is 2^M. This possibility was mentioned by Betts,[8] and we will only discuss it briefly. We can define $2(2^M - 1)$ direct product matrices of order 2^M in direct analogy with Eq. (3.4). The interaction energies between atoms on nearest-neighbor sites are chosen so that the 2-particle transfer matrix can be expanded in terms of the direct product matrices as in (3.5) and (3.10). It is also clear that the spinor algebras introduced in (3.17) can be generalized and lead to M sets of commuting spinors in this case. The dual transformation exists for all these models and takes the form given in (3.22).

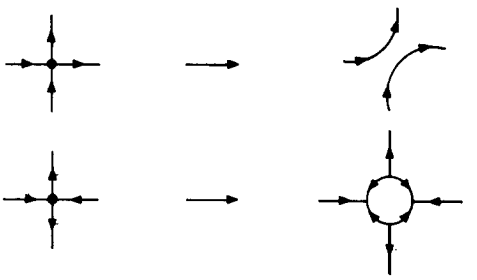

FIG. 3. Deformation of vertices of valence four.

ACKNOWLEDGMENTS

We would like to thank Professor L. Onsager for a valuable discussion of dual transformations. We are also indebted to Professor G. Benedek for hospitality at M.I.T., where this work was completed.

APPENDIX A

It is possible to deform the directed polygonal planar graphs to directed planar graphs. We simply suppress all vertices with only two edges. Now consider as one continuous directed edge a path which previously passed only through such suppressed vertices. We are left with a graph whose vertices contain only three or four edges. The number of edges at a vertex is called the vertex valence number ν. In our case $\nu = 3$ or 4. The vertices with $\nu = 4$ fall into two types represented in Fig. 3. Deform these two types as shown in Fig. 3.

We have now obtained a regular graph all of whose vertex numbers are equal; in this case $\nu = 3$. In a circuit enclosing only one face, an even number of edges are traversed. This is true because each time we pass through a vertex with $\nu = 3$, the direction is reversed. Since we must return to the original direction, we must traverse an even number of edges. The number of edges surrounding a given face is referred to as the face valence number. Ore[16] gives the following statement essentially as a corollary to a theorem: A regular graph of valence three is face colorable in three colors if and only if all faces have even valence. Thus all deformed graphs are three colorable. This clearly implies that all admissible diagrams are also three colorable. Any three coloration will yield either the directed graph or the graph with all directions reversed.

APPENDIX B

The algebra defined by Eqs. (2.31) and (2.32) has some interesting properties. All elements are units with vanishing trace. The multiplicative group defined by all products contains $3 \cdot 3^n \times 3^n$ traceless elements, whose cubes are the identity. Since $1 + \omega + \omega^2 = 0$, the algebra contains $2 \cdot 3^n \times 3^n$ linearly

independent elements over the reals or $3^n \times 3^n$ elements over the complexes. It is thus a basis for all $3^n \times 3^n$ matrices, in other words a complete matrix algebra. While this algebra satisfies no simple commutators or anticommutators, it satisfies the following relations:

$$\begin{aligned} Q_k\pi_l &= \omega\pi_l Q_k, \quad k \geq l, \\ \pi_k\pi_l &= \omega^{-1}\pi_l\pi_k, \quad k > l, \\ Q_k\pi_l &= \omega^{-1}\pi_l Q_k, \quad k < l, \\ Q_kQ_l &= \omega^{-1}Q_lQ_k, \quad k > l. \end{aligned}$$

Some of the triple terms such as $\pi_1\pi_2\pi_3$ equal $\pi_3\pi_2\pi_1$, but this is not a general property of the algebra. Thus, this algebra does not serve as a representation for parastatistics.

If we write the series $\pi_1, Q_1, \pi_2, Q_2, \pi_3, Q_3, \cdots, \pi_n, Q_n$, then, excluding the end factors, each entry has the same commutation relations with its neighbors; i.e., $\pi_1Q_1 = \omega^{-1}Q_1\pi_1$, $Q_1\pi_2 = \omega^{-1}\pi_2Q_1 \cdots$. This is essentially the mathematical structure behind the fact that the dual transformation is an automorphism. If we define a transformation partially by $D: \pi_1 \to Q_1 \to \pi_2 \to \cdots \pi_n \to Q_n \to \Phi(Q_n)$ such that $\Phi(U) = U^K$ and $D: \Lambda_1 A \to B \to \Lambda_1 A$, then $\Phi(U) = U^{-1}$ and $\Phi(Q_n) = \omega^{-1}\pi_1 U^{-2}$ complete the automorphism and preserve the commutation relations of Q_n with π_1 and π_n. The transformation given above serves the general q-component model. In the 3-component system, we have $\omega^{-1} = \omega^2$ and $U^{-2} = U$.

APPENDIX C

The critical energy E_c can be determined by assuming that the free energy is a differentiable function. This implies that f is differentiable with respect to its argument. With $kT\theta = 1$, we have at high- and low-temperatures, respectively,

$$E = -2\frac{\partial}{\partial\theta}\log u - \frac{f'(x/u)}{f(x/u)}\frac{\partial(x/u)}{\partial\theta}, \quad T \geq T_c,$$

$$E = -2\frac{\partial \log\alpha}{\partial\theta} - \frac{f'(\beta/\alpha)}{f(\beta/\alpha)}\frac{\partial(\beta/\alpha)}{\partial\theta}, \quad T \leq T_c.$$

The unknown function f'/f can be eliminated at the critical point to obtain

$$E_c = -\left[\frac{2\alpha'}{\alpha}\left(\frac{x}{u}\right)_c' - \frac{2u'}{u}\left(\frac{\beta}{\alpha}\right)_c'\right] \Big/ \left[\left(\frac{x}{u}\right)_c' - \left(\frac{\beta}{\alpha}\right)_c'\right].$$

The derivatives taken with respect to θ at the critical temperature satisfy $(x/u)_c' = -(\beta/\alpha)_c'$, so that

$$E_c = \epsilon_0 + \epsilon_1 - (\epsilon_1 - \epsilon_0)/\sqrt{q}.$$

* A preliminary report of this paper was published in the Bull. Am. Phys. Soc. **15**, 655 (1970).
† Supported in part by the National Science Foundation.

[1] An excellent review of the Ising model has been given by G. F. Newell and E. Montroll, Rev. Mod. Phys. **25**, 353 (1953).
[2] H. A. Kramers and G. H. Wannier, Phys. Rev. **60**, 252 (1941).
[3] L. Onsager, Phys. Rev. **65**, 117 (1944).
[4] B. Kaufman, Phys. Rev. **76**, 1232 (1949).
[5] J. Ashkin and E. Teller, Phys. Rev. **64**, 178 (1943).
[6] R. B. Potts, Proc. Cambridge Phil. Soc. **48**, 106 (1952).
[7] We are indebted to Professor M. E. Fisher for sending us a copy of his unpublished notes.
[8] D. D. Betts, Can. J. Phys. **42**, 1564 (1964).
[9] M. Suzuki, Progr. Theoret. Phys. (Kyoto) **37**, 770 (1967).
[10] O. Ore, *Theory of Graphs* (American Mathematical Society, Providence, R.I., 1962), Vol. 38.
[11] F. Harary, *Graph Theory* (Addison-Wesley, New York, 1969).
[12] R. J. Baxter, J. Math. Phys. **11**, 3116 (1970).
[13] E. H. Lieb, Phys. Rev. **162**, 162 (1967).
[14] The last factor is required to preserve the commutation relations of the π, Q. (See Appendix B.)
[15] R. Brauer and H. Weyl, Am. J. Math. **57**, 425 (1935).
[16] O. Ore, *The Four Color Problem* (Academic, New York, 1967), p. 227.

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 12, NUMBER 3 MARCH 1971

Effect of Molecular Collision Frequency on Solutions of the Linearized Boltzmann Equation*

THOMAS G. RICHARDSON AND LAWRENCE SIROVICH
Division of Applied Mathematics, Brown University, Providence, Rhode Island 02912

(Received 8 February 1970)

The linearized Boltzmann equation is considered for steady-state oscillations. Denoting molecular collision frequency by $\nu(\xi)$ and writing $\nu = O(\xi^\alpha)$ for ξ large, we show that solutions for $x \to \infty$ behave like $\exp(-x^{2/(3-\alpha)})$. This shows that the continuous spectra dominates hydrodynamics for all except the rigid sphere or radial cutoff case ($\alpha = 1$).

1. INTRODUCTION

It is a well-accepted fact and certainly no surprise that the behavior of a gas near a wall is described by kinetic theory equations and not simply by fluid mechanical ones (at least under most conditions). Several investigations[1-3] have indicated that flow far from a boundary is also a nonhydrodynamical regime. We are speaking, in particular, of the $O(\exp(-x^{2/3}))$ falloff at large distances predicted in sound propagation[2] and shock structure[1] for constant molecular collision frequencies and a somewhat more complicated but still nonhydrodynamical falloff for velocity

independent elements over the reals or $3^n \times 3^n$ elements over the complexes. It is thus a basis for all $3^n \times 3^n$ matrices, in other words a complete matrix algebra. While this algebra satisfies no simple commutators or anticommutators, it satisfies the following relations:

$$\begin{aligned} Q_k\pi_l &= \omega\pi_l Q_k, \quad k \geq l, \\ \pi_k\pi_l &= \omega^{-1}\pi_l\pi_k, \quad k > l, \\ Q_k\pi_l &= \omega^{-1}\pi_l Q_k, \quad k < l, \\ Q_kQ_l &= \omega^{-1}Q_lQ_k, \quad k > l. \end{aligned}$$

Some of the triple terms such as $\pi_1\pi_2\pi_3$ equal $\pi_3\pi_2\pi_1$, but this is not a general property of the algebra. Thus, this algebra does not serve as a representation for parastatistics.

If we write the series $\pi_1, Q_1, \pi_2, Q_2, \pi_3, Q_3, \cdots, \pi_n, Q_n$, then, excluding the end factors, each entry has the same commutation relations with its neighbors; i.e., $\pi_1Q_1 = \omega^{-1}Q_1\pi_1$, $Q_1\pi_2 = \omega^{-1}\pi_2Q_1 \cdots$. This is essentially the mathematical structure behind the fact that the dual transformation is an automorphism. If we define a transformation partially by $D: \pi_1 \to Q_1 \to \pi_2 \to \cdots \pi_n \to Q_n \to \Phi(Q_n)$ such that $\Phi(U) = U^K$ and $D: \Lambda_1 A \to B \to \Lambda_1 A$, then $\Phi(U) = U^{-1}$ and $\Phi(Q_n) = \omega^{-1}\pi_1 U^{-2}$ complete the automorphism and preserve the commutation relations of Q_n with π_1 and π_n. The transformation given above serves the general q-component model. In the 3-component system, we have $\omega^{-1} = \omega^2$ and $U^{-2} = U$.

APPENDIX C

The critical energy E_c can be determined by assuming that the free energy is a differentiable function. This implies that f is differentiable with respect to its argument. With $kT\theta = 1$, we have at high- and low-temperatures, respectively,

$$E = -2\frac{\partial}{\partial\theta}\log u - \frac{f'(x/u)}{f(x/u)}\frac{\partial(x/u)}{\partial\theta}, \quad T \geq T_c,$$

$$E = -2\frac{\partial \log\alpha}{\partial\theta} - \frac{f'(\beta/\alpha)}{f(\beta/\alpha)}\frac{\partial(\beta/\alpha)}{\partial\theta}, \quad T \leq T_c.$$

The unknown function f'/f can be eliminated at the critical point to obtain

$$E_c = -\left[\frac{2\alpha'}{\alpha}\left(\frac{x}{u}\right)_c' - \frac{2u'}{u}\left(\frac{\beta}{\alpha}\right)_c'\right] \Big/ \left[\left(\frac{x}{u}\right)_c' - \left(\frac{\beta}{\alpha}\right)_c'\right].$$

The derivatives taken with respect to θ at the critical temperature satisfy $(x/u)_c' = -(\beta/\alpha)_c'$, so that

$$E_c = \epsilon_0 + \epsilon_1 - (\epsilon_1 - \epsilon_0)/\sqrt{q}.$$

* A preliminary report of this paper was published in the Bull. Am. Phys. Soc. **15**, 655 (1970).
† Supported in part by the National Science Foundation.

[1] An excellent review of the Ising model has been given by G. F. Newell and E. Montroll, Rev. Mod. Phys. **25**, 353 (1953).
[2] H. A. Kramers and G. H. Wannier, Phys. Rev. **60**, 252 (1941).
[3] L. Onsager, Phys. Rev. **65**, 117 (1944).
[4] B. Kaufman, Phys. Rev. **76**, 1232 (1949).
[5] J. Ashkin and E. Teller, Phys. Rev. **64**, 178 (1943).
[6] R. B. Potts, Proc. Cambridge Phil. Soc. **48**, 106 (1952).
[7] We are indebted to Professor M. E. Fisher for sending us a copy of his unpublished notes.
[8] D. D. Betts, Can. J. Phys. **42**, 1564 (1964).
[9] M. Suzuki, Progr. Theoret. Phys. (Kyoto) **37**, 770 (1967).
[10] O. Ore, *Theory of Graphs* (American Mathematical Society, Providence, R.I., 1962), Vol. 38.
[11] F. Harary, *Graph Theory* (Addison-Wesley, New York, 1969).
[12] R. J. Baxter, J. Math. Phys. **11**, 3116 (1970).
[13] E. H. Lieb, Phys. Rev. **162**, 162 (1967).
[14] The last factor is required to preserve the commutation relations of the π, Q. (See Appendix B.)
[15] R. Brauer and H. Weyl, Am. J. Math. **57**, 425 (1935).
[16] O. Ore, *The Four Color Problem* (Academic, New York, 1967), p. 227.

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 12, NUMBER 3 MARCH 1971

Effect of Molecular Collision Frequency on Solutions of the Linearized Boltzmann Equation*

THOMAS G. RICHARDSON AND LAWRENCE SIROVICH
Division of Applied Mathematics, Brown University, Providence, Rhode Island 02912

(Received 8 February 1970)

The linearized Boltzmann equation is considered for steady-state oscillations. Denoting molecular collision frequency by $\nu(\xi)$ and writing $\nu = O(\xi^\alpha)$ for ξ large, we show that solutions for $x \to \infty$ behave like $\exp(-x^{2/(3-\alpha)})$. This shows that the continuous spectra dominates hydrodynamics for all except the rigid sphere or radial cutoff case ($\alpha = 1$).

1. INTRODUCTION

It is a well-accepted fact and certainly no surprise that the behavior of a gas near a wall is described by kinetic theory equations and not simply by fluid mechanical ones (at least under most conditions). Several investigations[1-3] have indicated that flow far from a boundary is also a nonhydrodynamical regime. We are speaking, in particular, of the $O(\exp(-x^{2/3}))$ falloff at large distances predicted in sound propagation[2] and shock structure[1] for constant molecular collision frequencies and a somewhat more complicated but still nonhydrodynamical falloff for velocity

dependent collision frequencies recently shown in Refs. 3 and 4.

The purpose of this investigation is to demonstrate the effect of intermolecular collision frequency on solutions of the linearized Boltzmann equation. We will carry out our calculations in connection with a problem in steady-state oscillations. However, the results we obtain should be equally applicable to other problems and to transport equations more general than the Boltzmann equation. In view of the somewhat heavy calculations, we give below an outline of our results and discuss them here rather than in the next section.

A first result is that the collision frequency $\nu(\xi)$ for high speed molecules has a direct effect on the nature of the solution. In fact, writing

$$\nu(\xi) = O(\xi^{\alpha}) \tag{1.1}$$

for ξ large (for noncharged particles $0 \leq \alpha \leq 1$), we find, for example, that the density ρ is given by

$$\rho = O(e^{-x^{2/(3-\alpha)}}). \tag{1.2}$$

This result has already been shown in an approximate method for certain kinetic models.[3,4] Our demonstration is given within the framework of the linearized Boltzmann equation itself.

For oscillations of frequency ω, hydrodynamics predicts that the falloff of oscillations at large distances is given by

$$\rho = O(e^{-k(\omega)x}) \tag{1.3}$$

(where $\operatorname{Re} k > 0$, $\omega \neq 0$). From this we see that hydrodynamic theory at large distances is small compared to the kinetic theory prediction except when $\alpha = 1$. Hence the region at infinity is a kinetic theory boundary layer.

The region near a wall is intuitively, at least, a free-flow regime; however, the region at infinity, although containing particles not having undergone collisions, is not a free-flow regime. We determine (1.2) by a stationary exponent calculation of an integral. Examination of this shows that, for a fixed but large value of x, the main contribution in the calculation comes from particles whose speed is $\xi = O(x^{1/(3-\alpha)})$. Now the free path of a particle moving with a speed ξ is ξ/ν; therefore, for fast particles, the free path $\lambda = O(\xi^{1-\alpha})$ by (1.1). Hence the main contribution to the evaluation (1.2) comes from particles having a free path

$$\lambda = O(x^{(1-\alpha)/(3-\alpha)}).$$

Therefore, these particles for all values of α have undergone many collisions.

This raises an interesting point with regard to the Chapman–Enskog procedure. There, it will be recalled, it is assumed that spatial derivatives are slowly varying with respect to the mean free path. It is, of course, clear that the procedure is not uniform in the velocity since the combination $\boldsymbol{\xi} \cdot \partial/\partial \mathbf{x}$ occurs. Our present investigation therefore demonstrates that, for $\alpha < 1$ and for certain regimes, $\boldsymbol{\xi}$ in this combination cannot be regarded as small.

As a next point, we take up the question of where classical hydrodynamics is valid. For low-frequency phenomena, $k(\omega)$ in (1.3) is $O(\omega^2)$. Comparing this with (1.2), we can say that in the low-frequency limit at least hydrodynamic theory is valid for

$$x \ll \omega^{-2[(3-\alpha)/(1-\alpha)]}.$$

Therefore, only for $\alpha = 1$ (effectively rigid sphere molecules) does the hydrodynamic region extend to infinity.

2. STEADY-STATE OSCILLATIONS

The problem of steady-state oscillations in a half-space has been discussed at length; the equation governing this is

$$\left(\nu(\xi) + i\omega + \xi_1 \frac{\partial}{\partial x}\right) g(x, \boldsymbol{\xi}) = Kg, \tag{2.1}$$

where ω is the frequency of oscillation and the linearized Boltzmann operator has been split into the difference $K - \nu$, ν being the molecular collision frequency. A discussion of the spectra of the operator $(i\omega + \xi_1(\partial/\partial x) + \nu - K)$ has been given in Ref. 5. In general, the spectra consists of point spectra and a two-dimensional region of continuous spectra. For (2.1), as is well known, only the outgoing distribution g ($\xi_1 > 0$) is in any way specified at $x = 0$. We will, however, not really consider any specific boundary-value problem. Formally we will regard the solution as known and then, from this, seek properties of it.

Regarding the right-hand side of (2.1) as known, we formally integrate and find

$$\begin{aligned} g(x, \boldsymbol{\xi}) = {} & H(\xi_1) g_0(\boldsymbol{\xi}) \exp\left(\frac{-(\nu + i\omega)x}{\xi_1}\right) \\ & + H(\xi_1) \int_0^x \frac{1}{\xi_1} (Kg)(s, \boldsymbol{\xi}) \\ & \times \exp\left(\frac{-(\nu + i\omega)(x - s)}{\xi_1}\right) ds \\ & - H(-\xi_1) \int_x^{\infty} \frac{1}{\xi_1} (Kg)(s, \boldsymbol{\xi}) \\ & \times \exp\left(\frac{-(\nu + i\omega)(x - s)}{\xi_1}\right) ds, \end{aligned} \tag{2.2}$$

where H is the Heaviside function and $g_o(\xi) = g(x = 0, \xi_1 > 0)$. Then, combining the last two terms of (2.2), we obtain

$$g(x, \xi) = H(\xi_1)g_o(\xi) \exp\left(\frac{-(\nu + i\omega)x}{\xi_1}\right) + \left(H(\xi_1)\int_0^x + H(-\xi_1)\int_0^\infty\right)\frac{1}{|\xi_1|}(Kg)(s, \xi) \times \exp\left(\frac{-(\nu + i\omega)|x - s|}{|\xi_1|}\right) ds \equiv (Bg_o)(x, \xi) + (WKg)(x, \xi), \tag{2.3}$$

where B and W represent the linear operators defined through (2.3). Note that (2.3) reduces the problem to one in integral equations; we will, however, not pursue this line of investigation here. We now solve for a moment of g, say the density

$$\rho(x) = (1, g)(x) \equiv \int_{-\infty}^{\infty} 1 \cdot g(x, \xi)(2\pi)^{-\frac{3}{2}} \exp(-\tfrac{1}{2}\xi^2)\, d\xi,$$

so that

$$\rho(x) = (1, Bg_o)(x) + (1, WKg)(x). \tag{2.4}$$

It is known that the term $(1, WKg)(x)$ contributes point spectra as well as continuous spectra to the solution. $(1, Bg_o)(x)$ contributes only continuous spectra. Here we examine the latter effect on $\rho(x)$ as x becomes large.

As will be clear, our discussion will apply to both normal and transverse oscillations, but for the sake of simplicity we assume purely normal oscillations in which case $g = g(\xi, \xi_1)$. Then we write the integral as

$$(1, Bg_o)(x) = \frac{1}{(2\pi)^{\frac{1}{2}}}\int_0^\infty\int_0^1 \xi^2 g_o(\xi, \mu) \times \exp\left(-\frac{\xi^2}{2} - \frac{[\nu(\xi) + i\omega]x}{\xi\mu}\right) d\xi\, d\mu,$$

where μ is the cosine of the polar angle and $\xi = |\xi|$. Considering first the μ integration, it is clear that the exponential term involving μ is a maximum when $\mu = 1$, since $c = [\nu(\xi) + i\omega]/\xi$ is such that

$$|\arg c| < \tfrac{1}{2}\pi - \delta,\ \delta > 0.$$

Under the transformation $\mu = 1/(1 + p)$ this part of the calculation is reduced to the Laplace integral

$$e^{-xc}\int_0^\infty e^{-xcp} g_o\left(\xi, \frac{1}{1 + p}\right)\frac{dp}{(1 + p)^2}.$$

Watson's lemma applies, giving the μ integral asymptotically equal to $(x \to \infty)$

$$e^{-xc} g_o(\xi, 1)[(xc)^{-1} + O(x^{-2}c^{-2})].$$

Thus, we are left with

$$(1, Bg_o)(x) \sim \frac{1}{(2\pi)^{\frac{1}{2}}}\int_0^\infty \frac{\xi^3 g_o(\xi, 1)}{x[\nu(\xi) + i\omega]} \times \exp\left(-\frac{\xi^2}{2} - \frac{[\nu(\xi) + i\omega]x}{\xi}\right) d\xi. \tag{2.5}$$

The stationary points of the exponential function

$$f(\xi; x) = -\tfrac{1}{2}\xi^2 - [\nu(\xi) + i\omega]x/\xi$$

are given by

$$-\xi_0 - \frac{\nu'(\xi)_0}{\xi_0}x + \frac{\nu(\xi)_0 + i\omega}{\xi^2_0}x = 0 \tag{2.6}$$

and the second derivative is

$$-f_{\xi\xi}(\xi_0; x) = 3 + \frac{\nu''(\xi_0)}{\xi_0}x. \tag{2.7}$$

Take ξ to be a complex variable and expand $g_o(\xi, 1)$ about the solution ξ_o of (2.6); then, with (2.7) to lowest order, the method of steepest descent gives

$$(1, Bg_o) \sim \frac{e^{f(\xi_o; x)}}{[-f_{\xi\xi}(\xi_o; x)]^{\frac{1}{2}}}\frac{g_o(\xi_o, 1)\xi_o^3}{x[\nu(\xi_o) + i\omega]}. \tag{2.8}$$

To make (2.8) specific, we suppose that, for ξ large, ν has the asymptotic expansion

$$\nu(\xi) = k\xi^\alpha + \alpha V + O(\xi^{-\beta}),$$

with $0 \le \alpha \le 1$ and k, V, and β positive constants. The case $\alpha = 1$ corresponds to rigid sphere molecules or radial cutoff. $\alpha = 0$ gives a constant collision frequency as in the Krook model equation. Then the solution of (2.6) is, for x large, given by

$$\begin{aligned} \xi_o &= [(k + i\omega)x]^{\frac{1}{3}} + O(x^{(1-\beta)/3}), && \alpha = 0, \\ &= [(1 - \alpha)kx]^{1/(3-\alpha)} + O(x^{(1-\alpha)/(3-\alpha)}), && 0 < \alpha < 1, \\ &= [(V + i\omega)x]^{\frac{1}{3}} + O(x^{(1-\beta)/3}), && \alpha = 1. \end{aligned} \tag{2.9}$$

Then for (2.7) we have

$$\begin{aligned} f_{\xi\xi}(\xi_o; x) &= -3 + O(x^{-\beta/3}), && \alpha = 0, \\ &= -3 + \alpha + O(x^{-\alpha/(3-\alpha)}), && 0 < \alpha < 1, \\ &= -3 + O(x^{-\beta/3}), && \alpha = 1 \end{aligned} \tag{2.10}$$

and

$$\begin{aligned} f(\xi_o; x) &= -\tfrac{3}{2}[(k + i\omega)x]^{\frac{2}{3}} + O(x^{(2-\beta)/3}), && \alpha = 0, \\ &= -\tfrac{1}{2}(3 - \alpha)(1 - \alpha)^{(\alpha-1)/(3-\alpha)}(kx)^{2/(3-\alpha)} \\ &\qquad + O(x^{(2-\alpha)/(3-\alpha)}), && 0 < \alpha < 1, \\ &= -(kx) + O(x^{\frac{2}{3}}), && \alpha = 1. \end{aligned} \tag{2.11}$$

$f(\xi_o; x)$ can be written compactly as

$$f(\xi_o; x) = -[kx + i\omega x\delta_{\alpha 0}]^{2/(3-\alpha)} \times \tfrac{1}{2}(3-\alpha)(1-\alpha)^{(\alpha-1)/(3-\alpha)} + o(x^{2/(3-\alpha)}), \quad 0 \leq \alpha \leq 1, \quad (2.12)$$

where the order term is more precisely given by (2.11). It can be shown by using (2.9)–(2.12) that paths in the complex plane can always be found such that the method of steepest descent is valid, and that (2.8) is then given as

$$(1, Bg_o) \sim \frac{g_o(\xi_o, 1)}{[3-\alpha+\delta_{\alpha 1}]^{\frac{1}{2}}}\left[1-\alpha+\delta_{\alpha 1}\frac{(V+i\omega)^{\frac{2}{3}}}{kx^{\frac{1}{3}}}\right] \times \exp\{[kx + i\omega x\delta_{\alpha 0}]^{2/(3-\alpha)} \times \tfrac{1}{2}(3-\alpha)(1-\alpha)^{(\alpha-1)/(3-\alpha)} + o(x^{2/(3-\alpha)})\}. \quad (2.13)$$

where $\delta_{\alpha\beta}$ is the Kronecker delta.

Thus we observe that $(1, Bg_o)(x)$ behaves essentially as $\exp(-ax^{2/(3-\alpha)})$, $a > 0$, for x large. This result, for $\alpha = 0$, was first obtained in Refs. 1 and 2 using the Krook equation. For $\alpha > 0$ the result was found in Refs. 3 and 4 by using an approximate method for kinetic models. Here we have demonstrated that this result holds in general for the linearized Boltzmann equation.

* The results given in this paper were found in the course of research sponsored by the Office of Naval Research.

[1] G. Ya. Lyubarskii, Zh. Eksp. Teor. Fiz. **40**, 1050 (1961) [Sov. Phys. JETP **13**, 740 (1961)].
[2] H. S. Ostrowsky and D. J. Kleitman, Nuovo Cimento **44B**, 49 (1966).
[3] L. Sirovich and H. York, Phys. Fluids **13**, 1731 (1970).
[4] L. Sirovich and H. York, "The Rayleigh Problem and Transversal Oscillations in Kinetic Theory," Brown University, Division of Applied Mathematics Technical Report, 1970.
[5] L. Sirovich and J. K. Thurber, J. Math. Phys. **10**, 239 (1969).

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 12, NUMBER 3 MARCH 1971

Stochastic Theory of Quantum Mechanics for Particles with Spin

L. DE LA PEÑA-AUERBACH
Instituto de Fisica, Universidad Nacional de Mexico, Mexico, D.F.

(Received 22 May 1970)

The stochastic theory of quantum mechanics is further developed to include the problem of extended rigid particles, thus allowing the introduction of spin. It is demonstrated that the stochastic equations for the system's center of mass give rise to a generalized Schrödinger equation for integral or half-integral spin; in the particular case of spin $\frac{1}{2}$, upon elimination of the internal variables, Pauli's equation is obtained. A formal simplified relativistic extension of the theory is worked out and shown to lead to Dirac's equation in the case of spin $\frac{1}{2}$ and for a gyromagnetic ratio equal to 2; in the case of arbitrary spin, the theory gives an equation of the Feynman–Gell-Mann type.

I. INTRODUCTION

In a series of papers[1] we have proposed an elementary theory for a classical particle subject to a random interaction with its surroundings. In (I) it was shown that such a stochastic theory contains as a particularly simple case the quantum mechanics of a nonrelativistic spinless particle under the action of an external potential. In (II), (III), and (IV) we demonstrated that the theory applies also to more general situations, as, for example, the electromagnetic case, and to a system of interacting particles; in particular, the two-body problem was studied more closely. The aim of this paper is to extend the theory to particles with spin. With this purpose and following the line of thought presented in the aforementioned papers, we shall consider our stochastic particle as a spinning rigid body. We are aware of the fact that such a model is not fashionable, due to its inherent difficulties,[2] and that more abstract and formal procedures, which assign no classical analog to the spin variable, are being preferred. Nevertheless, the introduction of spinning rigid bodies has given lately a series of interesting results, both in the study of electron spin in particular[3,4] and in connection with some attempts to understand the nature of the quantum numbers of elementary particles; two representative examples of such attempts are given in Refs. 5 and 6.

To start with, we consider a system of stochastic particles and apply to them the methods previously developed. Upon introduction of the constraints defining a rotating rigid body, the theory gives a Schrödinger equation, except for some additional terms due both to the particle's extension and to its spin. The amplitude ψ is now a function of the center-of-mass

$f(\xi_o; x)$ can be written compactly as

$$f(\xi_o; x) = -[kx + i\omega x\delta_{\alpha 0}]^{2/(3-\alpha)} \times \tfrac{1}{2}(3-\alpha)(1-\alpha)^{(\alpha-1)/(3-\alpha)} + o(x^{2/(3-\alpha)}), \quad 0 \leq \alpha \leq 1, \quad (2.12)$$

where the order term is more precisely given by (2.11). It can be shown by using (2.9)–(2.12) that paths in the complex plane can always be found such that the method of steepest descent is valid, and that (2.8) is then given as

$$(1, Bg_o) \sim \frac{g_o(\xi_o, 1)}{[3-\alpha+\delta_{\alpha 1}]^{\frac{1}{2}}}\left[1-\alpha+\delta_{\alpha 1}\frac{(V+i\omega)^{\frac{2}{3}}}{kx^{\frac{1}{3}}}\right] \times \exp\{[kx + i\omega x\delta_{\alpha 0}]^{2/(3-\alpha)} \times \tfrac{1}{2}(3-\alpha)(1-\alpha)^{(\alpha-1)/(3-\alpha)} + o(x^{2/(3-\alpha)})\}. \quad (2.13)$$

where $\delta_{\alpha\beta}$ is the Kronecker delta.

Thus we observe that $(1, Bg_o)(x)$ behaves essentially as $\exp(-ax^{2/(3-\alpha)})$, $a > 0$, for x large. This result, for $\alpha = 0$, was first obtained in Refs. 1 and 2 using the Krook equation. For $\alpha > 0$ the result was found in Refs. 3 and 4 by using an approximate method for kinetic models. Here we have demonstrated that this result holds in general for the linearized Boltzmann equation.

* The results given in this paper were found in the course of research sponsored by the Office of Naval Research.

[1] G. Ya. Lyubarskii, Zh. Eksp. Teor. Fiz. **40**, 1050 (1961) [Sov. Phys. JETP **13**, 740 (1961)].
[2] H. S. Ostrowsky and D. J. Kleitman, Nuovo Cimento **44B**, 49 (1966).
[3] L. Sirovich and H. York, Phys. Fluids **13**, 1731 (1970).
[4] L. Sirovich and H. York, "The Rayleigh Problem and Transversal Oscillations in Kinetic Theory," Brown University, Division of Applied Mathematics Technical Report, 1970.
[5] L. Sirovich and J. K. Thurber, J. Math. Phys. **10**, 239 (1969).

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 12, NUMBER 3 MARCH 1971

Stochastic Theory of Quantum Mechanics for Particles with Spin

L. DE LA PEÑA-AUERBACH
Instituto de Fisica, Universidad Nacional de Mexico, Mexico, D.F.

(Received 22 May 1970)

The stochastic theory of quantum mechanics is further developed to include the problem of extended rigid particles, thus allowing the introduction of spin. It is demonstrated that the stochastic equations for the system's center of mass give rise to a generalized Schrödinger equation for integral or half-integral spin; in the particular case of spin $\frac{1}{2}$, upon elimination of the internal variables, Pauli's equation is obtained. A formal simplified relativistic extension of the theory is worked out and shown to lead to Dirac's equation in the case of spin $\frac{1}{2}$ and for a gyromagnetic ratio equal to 2; in the case of arbitrary spin, the theory gives an equation of the Feynman–Gell-Mann type.

I. INTRODUCTION

In a series of papers[1] we have proposed an elementary theory for a classical particle subject to a random interaction with its surroundings. In (I) it was shown that such a stochastic theory contains as a particularly simple case the quantum mechanics of a nonrelativistic spinless particle under the action of an external potential. In (II), (III), and (IV) we demonstrated that the theory applies also to more general situations, as, for example, the electromagnetic case, and to a system of interacting particles; in particular, the two-body problem was studied more closely. The aim of this paper is to extend the theory to particles with spin. With this purpose and following the line of thought presented in the aforementioned papers, we shall consider our stochastic particle as a spinning rigid body. We are aware of the fact that such a model is not fashionable, due to its inherent difficulties,[2] and that more abstract and formal procedures, which assign no classical analog to the spin variable, are being preferred. Nevertheless, the introduction of spinning rigid bodies has given lately a series of interesting results, both in the study of electron spin in particular[3,4] and in connection with some attempts to understand the nature of the quantum numbers of elementary particles; two representative examples of such attempts are given in Refs. 5 and 6.

To start with, we consider a system of stochastic particles and apply to them the methods previously developed. Upon introduction of the constraints defining a rotating rigid body, the theory gives a Schrödinger equation, except for some additional terms due both to the particle's extension and to its spin. The amplitude ψ is now a function of the center-of-mass

(c.m.) coordinates and of the internal degrees of freedom, e.g., the Euler angles. From usual considerations it follows that the spin is quantized and may take integral or half-integral values. When separating c.m. from rotational coordinates, we obtain, for spin $\frac{1}{2}$, just Pauli's equation—ignoring some minor corrections due to the particle's finite radius, a point which will be briefly commented on in the text.

In the last part of the paper, we make a preliminary investigation on the possibility of extending the theory to the relativistic case. This may be achieved by first constructing the stochastic theory of a relativistic point particle and then trying to introduce the additional internal degrees of freedom. The first part may be accomplished following the procedure applied to the nonrelativistic case, the results being, as one may expect in advance, the covariant generalization of the nonrelativistic equations,[7] a result which we take for granted in this paper. The second step is much more difficult because classical relativistic theories of extended bodies are still under discussion,[5,6,8] and hence it is not a simple job to select one among the many (occasionally contradictory or at least nonequivalent) existing theories. Here we get around this difficulty by the simple and common expedient[8] of writing the formal covariant generalization of the nonrelativistic spin terms previously obtained. The relativistic stochastic equations constructed in this form lead us, upon integration, to a "wave equation" of the Feynman–Gell-Mann type,[9] but written in terms of relativistic internal variables. To obtain more familiar results and at the same time maintain things as simple as possible, we go over to the case of spin $\frac{1}{2}$, making some approximations which allow us to use the nonrelativistic theory; known procedures lead us then to the corresponding first-order equation, which turns out to be in this particular case just Dirac's equation if the gyromagnetic ratio is taken equal to two.

II. STOCHASTIC EQUATIONS FOR THE SPINNING RIGID BODY

Consider a system of particles subject both to a stochastic interaction with the surrounding medium and to "external" interactions, i.e., interactions between themselves and with given external fields. According to the theory developed in (I), we may write for particle α (Greek indices are used to label the particles)

$$\begin{aligned} \mathfrak{D}_c m^\alpha \mathbf{v}^\alpha - \mathfrak{D}_s m^\alpha \mathbf{u}^\alpha &= \mathbf{f}_0^{(+)\alpha}, \\ \mathfrak{D}_s m^\alpha \mathbf{v}^\alpha + \mathfrak{D}_c m^\alpha \mathbf{u}^\alpha &= \mathbf{f}_0^{(-)\alpha}; \end{aligned} \tag{1}$$

m^α is the mass of the particle with systematic and stochastic velocities $\mathbf{v}^\alpha$ and $\mathbf{u}^\alpha$, respectively; $\mathbf{f}_0^{(\pm)\alpha}$ stands for the external force acting on particle α, the plus or minus sign referring to its behavior under time-reversal [see (III)].

From (I) and (IV) we also know that

$$\begin{aligned} \mathfrak{D}_c \mathbf{r}^\alpha &= \mathbf{v}^\alpha, \\ \mathfrak{D}_s \mathbf{r}^\alpha &= \mathbf{u}^\alpha. \end{aligned} \tag{2}$$

To second order, the operators $\mathfrak{D}_c$ and $\mathfrak{D}_s$ may be written schematically as

$$\begin{aligned} \mathfrak{D}_c &= \frac{\partial}{\partial t} + \sum_i v_i \frac{\partial}{\partial x_i}, \\ \mathfrak{D}_s &= \sum_i u_i \frac{\partial}{\partial x_i} + D \sum_i \frac{\partial}{\partial x_i} \frac{\partial}{\partial x_i}, \end{aligned} \tag{3}$$

the summation extending over all degrees of freedom.[10]

In order to constrain our system of particles to define a rotating rigid body, we write $\mathbf{r}^\alpha$ in the form

$$\mathbf{r}^\alpha = \mathbf{r} + \boldsymbol{\rho}^\alpha, \tag{4a}$$

where $\mathbf{r}$ is the c.m. coordinate

$$\mathbf{r} = \sum_\alpha \lambda^\alpha \mathbf{r}^\alpha \tag{4b}$$

and

$$m = \sum_\alpha m^\alpha, \quad m^\alpha = m\lambda^\alpha, \tag{4c}$$

m being the total mass of the system. Hence $\boldsymbol{\rho}^\alpha$, the position vector of particle α relative to the c.m., is a rotating vector of constant magnitude. From Eqs. (4a) and (4b) it follows trivially that

$$\sum_\alpha \lambda^\alpha \boldsymbol{\rho}^\alpha = 0. \tag{4d}$$

Fixing now a system of axes in the body, we see that it will rotate with given angular velocity $\boldsymbol{\Omega}$ with respect to a system of coordinates fixed in space. Hence, the total velocity of particle α may be written as

$$\mathbf{c}^\alpha = \mathbf{c} + \boldsymbol{\Omega} \times \boldsymbol{\rho}^\alpha, \tag{5a}$$

where, as usual,

$$\mathbf{c}^\alpha = \mathbf{v}^\alpha + \mathbf{u}^\alpha.$$

Separating systematic and stochastic components, we get

$$\begin{aligned} \mathbf{v}^\alpha &= \mathbf{v} + \boldsymbol{\omega} \times \boldsymbol{\rho}^\alpha = \mathfrak{D}_c \mathbf{r}^\alpha, \\ \mathbf{u}^\alpha &= \mathbf{u} + \boldsymbol{\eta} \times \boldsymbol{\rho}^\alpha = \mathfrak{D}_s \mathbf{r}^\alpha, \end{aligned} \tag{5b}$$

with

$$\mathbf{c} = \mathbf{v} + \mathbf{u}$$

and

$$\boldsymbol{\Omega} = \boldsymbol{\omega} + \boldsymbol{\eta}. \tag{5c}$$

Clearly, $\mathbf{v}$ and $\mathbf{u}$ stand for the systematic and stochastic velocities of the c.m., whereas $\boldsymbol{\omega}$ and $\boldsymbol{\eta}$ represent the systematic and stochastic components of the angular

velocity of rotation, respectively. From Eqs. (4a) and (5b) we have that

$$\mathbf{v} + \boldsymbol{\omega} \times \boldsymbol{\rho}^\alpha = \mathfrak{D}_c \mathbf{r} + \mathfrak{D}_c \boldsymbol{\rho}^\alpha,$$
$$\mathbf{u} + \boldsymbol{\eta} \times \boldsymbol{\rho}^\alpha = \mathfrak{D}_s \mathbf{r} + \mathfrak{D}_s \boldsymbol{\rho}^\alpha.$$

For $\boldsymbol{\rho}^\alpha = 0$, this set of equations reduces to

$$\mathbf{v} = \mathfrak{D}_c \mathbf{r}$$
$$\mathbf{u} = \mathfrak{D}_s \mathbf{r}. \tag{6a}$$

In particular, the position of the c.m. is defined just by $\boldsymbol{\rho}^\alpha = 0$; hence, Eqs. (6a) describe the motion of the c.m. Introducing Eqs. (6a) into the preceding set of equations, we obtain for the relative motion

$$\mathfrak{D}_c \boldsymbol{\rho}^\alpha = \boldsymbol{\omega} \times \boldsymbol{\rho}^\alpha,$$
$$\mathfrak{D}_s \boldsymbol{\rho}^\alpha = \boldsymbol{\eta} \times \boldsymbol{\rho}^\alpha. \tag{6b}$$

Clearly, in the Newtonian limit ($D \to 0$, $\mathbf{u} \to 0$) Eqs. (6) reduce correctly to the classical equations, i.e., $\dot{\mathbf{r}} = \mathbf{v}$, $\dot{\boldsymbol{\rho}}^\alpha = \mathbf{w} \times \boldsymbol{\rho}^\alpha$.

The total momentum of the system is

$$\mathbf{p} = \sum_\alpha m^\alpha \mathbf{c}^\alpha = m\mathbf{c} = m\mathbf{v} + m\mathbf{u} \tag{7}$$

and the total angular momentum of the body may be written as

$$\mathbf{j} = \sum_\alpha \mathbf{r}^\alpha \times \mathbf{p}^\alpha = \mathbf{r} \times \mathbf{p} + \mathbb{I} \cdot \boldsymbol{\Omega} = \mathbf{l} + \mathbf{s}, \tag{8}$$

$\mathbf{l}$ being the orbital angular momentum and $\mathbf{s}$ the spin; $\mathbb{I}$ stands for the inertia tensor with components

$$\mathbb{I}_{ij} = \sum_\alpha m^\alpha [(\boldsymbol{\rho}^\alpha)^2 \delta_{ij} - \rho_i^\alpha \rho_j^\alpha].$$

For simplicity, we shall consider a spherically symmetric, not necessarily homogeneous, mass distribution. Hence, we write

$$\mathbb{I}_{ij} = I\delta_{ij} \tag{9a}$$

with

$$I = \tfrac{2}{5} m a_2^2. \tag{9b}$$

The parameter a_2 is proportional to the radius of the mass distribution, the constant of proportionality depending on the particular form of such distribution; for any "reasonable" law we may expect its value to be of order unity.

We see from Eq. (8) that both the orbital and spin angular momenta have systematic and stochastic components, i.e.,

$$\mathbb{I} = \mathbb{I}_c + \mathbb{I}_s = m\mathbf{r} \times \mathbf{v} + m\mathbf{r} \times \mathbf{u} \tag{10a}$$

and

$$\mathbf{s} = \mathbf{s}_c + \mathbf{s}_s = I\boldsymbol{\Omega} \tag{10b}$$

with

$$\mathbf{s}_c = I\boldsymbol{\omega}, \quad \mathbf{s}_s = I\boldsymbol{\eta}.$$

Since the degrees of freedom have been reduced to six, Eqs. (3) take the form

$$\mathfrak{D}_c = \frac{\partial}{\partial t} + \mathbf{v} \cdot \nabla + \boldsymbol{\omega} \cdot \nabla_\xi,$$
$$\mathfrak{D}_s = \mathbf{u} \cdot \nabla + D\nabla^2 + \boldsymbol{\eta} \cdot \nabla_\xi + D_\xi \nabla_\xi^2, \tag{11}$$

where ∇ represents derivation with respect to the c.m. coordinates and ∇_ξ with respect to the internal variables and may be written, for example, in terms of Euler angles[11]; further,

$$D = \hbar/2m, \tag{12a}$$
$$D_\xi = \hbar/2I. \tag{12b}$$

The first of these relations has been obtained in several papers, in particular in (I). The second relation can be obtained, for example, by calculating the value of the auxiliary quantity

$$\Delta = \sum_{k,\alpha}{}' m^\alpha E\left\{\frac{\delta\rho_k^\alpha \delta\rho_k^\alpha}{2\Delta t}\right\} \tag{13}$$

as $\Delta t \to 0$; $E\{\ \}$ denotes mean value as defined in (I), and the prime is being used to indicate that the rigid-body constraints must be taken into account when performing the summation. The calculation will be carried out in two different ways: If we introduce the rigid-body constraints by writing

$$\delta\rho_k^\alpha = \sum_{i,j} \epsilon_{ijk} \delta\alpha_i \rho_j^\alpha,$$

where the $\delta\alpha_i$ represent the angular deviations occurring in Δt, we obtain with the aid of Eq. (9a) the value $\Delta = 3ID_\xi$. Here we have written

$$\lim_{\Delta t \to 0} E\left\{\frac{\delta\alpha_i \delta\alpha_j}{2\Delta t}\right\} = D_\xi \delta_{ij}.$$

The diagonal character of this matrix is self-evident, since $\delta\alpha_i$ and $\delta\alpha_j$, $i \neq j$, are stochastically independent; furthermore, from isotropy considerations it follows that the diagonal elements are equal. On the other hand, since $\delta\rho^\alpha$ is small, we may treat it as the translation of particle α as viewed from the c.m.; in this case the constraints are taken into account by considering only one particle along the k axis, i.e., by reducing the number of internal degrees of freedom to three. Hence, since all three directions are equivalent, we obtain[1] $\Delta = \sum_k m^\alpha \hbar/2m^\alpha = 3\hbar/2$. Equation (12b) is an immediate consequence of these two results.

Note that the operators $\mathfrak{D}_c$ and $\mathfrak{D}_s$ commute with $\sum_\alpha$, a property that will be frequently used in what follows.

The translational as well as the rotational equations of motion may be established by following a procedure similar to that used in the classical treatment of

the rigid body. Addition of Eqs. (1) over all particles leads to the translational equations

$$\mathfrak{D}_c m\mathbf{v} - \mathfrak{D}_s m\mathbf{u} = \mathbf{f}_0^{(+)},$$
$$\mathfrak{D}_s m\mathbf{v} + \mathfrak{D}_c m\mathbf{u} = \mathbf{f}_0^{(-)}, \tag{14}$$

where

$$\mathbf{f}_0^{(\pm)} = \sum_\alpha \mathbf{f}_0^{(\pm)\alpha} \tag{15}$$

represents the external forces acting on the c.m. The rotational equations of motion, which are obtained by taking the cross product of each of Eqs. (1) with $\mathbf{r}^\alpha$ and summing over all particles, are omitted since their derivation is rather lengthy and they are not relevant to our present purposes.

III. THE ELECTROMAGNETIC CASE

In this section we calculate $\mathbf{f}_0^{(\pm)}$ for the electromagnetic case. To achieve this, we write for the force acting on particle α, to second order,

$$\mathbf{f}_0^{(+)\alpha} = e^\alpha[\mathbf{E}^\alpha + (1/c)\mathbf{v}^\alpha \times \mathbf{H}^\alpha],$$
$$\mathbf{f}_0^{(-)\alpha} = e^\alpha[(D/c)\nabla \times \mathbf{H}^\alpha + (1/c)\mathbf{u}^\alpha \times \mathbf{H}^\alpha], \tag{16}$$

where the gauge $\nabla \cdot \mathbf{A} = 0$ has been used. Here $\mathbf{E}^\alpha$ and $\mathbf{H}^\alpha$ represent the external electric and magnetic field intensities at the point $\mathbf{r}^\alpha$ occupied by particle α. Equations (16) applied to a single charged particle lead us to Schrödinger's equation, as has been shown in (III), but it seems desirable to justify them at this point. Possibly the simplest way to derive them goes as follows. In (I) we have seen that, to go over from classical mechanics to the stochastic theory, the substitution

$$\mathbf{v} \to \mathbf{c} = \mathbf{v} + \mathbf{u},$$
$$\frac{d}{dt} \to \mathfrak{D} = \mathfrak{D}_c + \mathfrak{D}_s$$

must be made. Applying this recipe to the classical formula

$$\mathbf{f}_0 = -\frac{e}{c}\frac{d\mathbf{A}}{dt} + \frac{e}{c}(\mathbf{v}\cdot\nabla)\mathbf{A} + \frac{e}{c}\mathbf{v}\times(\nabla\times\mathbf{A}) - e\nabla\phi,$$

we obtain

$$\mathbf{f}_0 = \mathbf{f}_0^{(+)} + \mathbf{f}_0^{(-)} = -(e/c)\mathfrak{D}\mathbf{A} + (e/c)(\mathbf{v}+\mathbf{u})\cdot\nabla\mathbf{A} + (e/c)(\mathbf{v}+\mathbf{u})\times(\nabla\times\mathbf{A}) - e\nabla\phi.$$

Hence, separating terms with different behavior under time reversal,

$$\mathbf{f}_0^{(+)} = (-e/c)\mathfrak{D}_c\mathbf{A} + (e/c)(\mathbf{v}\cdot\nabla)\mathbf{A} + (e/c)\mathbf{v}\times(\nabla\times\mathbf{A}) - e\nabla\phi,$$
$$\mathbf{f}_0^{(-)} = (-e/c)\mathfrak{D}_s\mathbf{A} + (e/c)(\mathbf{u}\cdot\nabla)\mathbf{A} + (e/c)\mathbf{u}\times(\nabla\times\mathbf{A}). \tag{16'}$$

Equations (16) are obtained by developing $\mathfrak{D}_c$ and $\mathfrak{D}_s$ in Eq. (16') to second order, with

$$\mathbf{E} = -\frac{1}{c}\frac{\partial\mathbf{A}}{\partial t} - \nabla\phi \quad \text{and} \quad \mathbf{H} = \nabla\times\mathbf{A}.$$

For the calculation of $\mathbf{f}_0^{(\pm)}$ we expand $\mathbf{E}^\alpha$ and $\mathbf{H}^\alpha$ in a Taylor series about the c.m. and, to take into account possible effects due to the distribution of electricity in the volume occupied by the extended particle, we retain terms up to second order. Clearly, in analogy with Eqs. (4c) and (4d), we may write

$$\sum_\alpha e^\alpha = e, \tag{17a}$$

$$\sum_\alpha e^\alpha\rho^\alpha = 0. \tag{17b}$$

However, since the charge distribution does not necessarily coincide with the mass distribution, we introduce the quantity

$$g^\alpha = g(\rho^\alpha) = (e^\alpha/e)(m^\alpha/m)^{-1} \tag{18a}$$

and write

$$\sum_\alpha e^\alpha\rho_i^\alpha\rho_j^\alpha = \frac{e}{m}\sum_\alpha g^\alpha m^\alpha\rho_i^\alpha\rho_j^\alpha.$$

Hence, defining an average g by

$$g = \sum_\alpha g^\alpha m^\alpha(\rho^\alpha)^2 \Big/ \sum_\alpha m^\alpha(\rho^\alpha)^2, \tag{18b}$$

we may write for a spherically symmetric charge distribution, taking into account Eq. (9),

$$\sum_\alpha e^\alpha\rho_i^\alpha\rho_j^\alpha = \tfrac{1}{3}gea_2^2\delta_{ij}. \tag{19}$$

Using Eq. (16) to calculate $\mathbf{f}_0^{(\pm)}$, we obtain therefore

$$\mathbf{f}_0^{(+)} = e[\mathbf{E} + (1/c)\mathbf{v}\times\mathbf{H}] + (ge/2mc)[\nabla(\mathbf{s}_c\cdot\mathbf{H}) - (\mathbf{H}\cdot\nabla)\mathbf{s}_c - \mathbf{H}\times(\nabla\times\mathbf{s}_c)] + \tfrac{1}{6}gea_2^2[\nabla^2\mathbf{E} + (1/c)\mathbf{v}\times\nabla^2\mathbf{H}] \tag{20a}$$

and

$$\mathbf{f}_0^{(-)} = e[(D/c)\nabla\times\mathbf{H} + (1/c)\mathbf{u}\times\mathbf{H}] + (ge/2mc)[\nabla(\mathbf{s}_s\cdot\mathbf{H}) - (\mathbf{H}\cdot\nabla)\mathbf{s}_s - \mathbf{H}\times(\nabla\times\mathbf{s}_s)] + \tfrac{1}{6}gea_2^2[(D/c)\nabla^2(\nabla\times\mathbf{H}) + (1/c)\mathbf{u}\times\nabla^2\mathbf{H}]. \tag{20b}$$

The force $\mathbf{f}_0^{(+)}$ is classical, i.e., in the Newtonian limit it is given by the same Eq. (20a); on the other hand, $\mathbf{f}_0^{(-)}$ goes to zero in the Newtonian limit, which indicates its purely stochastic origin.

IV. THE SCHRÖDINGER EQUATION FOR ARBITRARY SPIN

As was shown in (I) and (III) for the case of spinless point particles, the first integral of the system of Eqs. (14) leads to Schrödinger's equation. Essentially the same procedure can be applied to the spinning extended particle. For this purpose, it is convenient to first rewrite Eqs. (14) in a more simple and suggestive form, by expressing them in terms of the following complex quantities[12]:

$$\begin{aligned} \mathbf{v}_q &= \mathbf{v} - i\mathbf{u}, \\ \mathbf{s}_q &= \mathbf{s}_c - i\mathbf{s}_s, \\ \mathbf{f}_q &= \mathbf{f}_0^{(+)} - i\mathbf{f}_0^{(-)}, \\ \mathfrak{D}_q &= \mathfrak{D}_c - i\mathfrak{D}_s. \end{aligned} \tag{21}$$

Using Eqs. (21) and their complex conjugates, we may write Eqs. (14) in the following form:

$$\mathfrak{D}_q m\mathbf{v}_q = \mathbf{f}_q, \tag{22a}$$

$$\mathfrak{D}_q^* m\mathbf{v}_q^* = \mathbf{f}_q^*. \tag{22b}$$

Owing to the symmetry of these equations, we may work with only one of them; we choose the first one and consider it a complex form of Newton's second law for stochastic particles.

In the electromagnetic case we have, according to Eqs. (20),

$$\mathbf{f}_q = \mathbf{f}_q^{(1)} + \mathbf{f}_q^{(2)} + \mathbf{f}_q^{(3)}, \tag{23a}$$

with

$$\mathbf{f}_q^{(1)} = e[\mathbf{E} - (iD/c)\nabla \times \mathbf{H} + (1/c)\mathbf{v}_q \times \mathbf{H}], \tag{23b}$$

$$\mathbf{f}_q^{(2)} = (ge/2mc)[\nabla(\mathbf{s}_q \cdot \mathbf{H}) - (\mathbf{H} \cdot \nabla)\mathbf{s}_q - \mathbf{H} \times (\nabla \times \mathbf{s}_q)], \tag{23c}$$

$$\mathbf{f}_q^{(3)} = \tfrac{1}{6}gea_2^2[\nabla^2\mathbf{E} - (iD/c)\nabla \times (\nabla^2\mathbf{H}) + (1/c)\mathbf{v}_q \times \nabla^2\mathbf{H}]; \tag{23d}$$

$\mathbf{f}_q^{(1)}$ represents a kind of complex force acting on the c.m.; $\mathbf{f}_q^{(2)}$ is due to the interaction of the spin with the magnetic field and, finally, $\mathbf{f}_q^{(3)}$ is a residual force due to the structure of the particle.

In general, Eqs. (22) and (23) together with the rotational equations of motion form an extremely complicated coupled system; we shall therefore restrict ourselves in this paper to a simple but relevant case, namely, when $\mathbf{H}$ is such that $\mathbf{s}_q$ does not depend on the c.m. coordinates and precesses around $\mathbf{H}$. This restriction is not essential from the point of view of the ideas proposed in the present paper, but represents a considerable simplification of the mathematical process. In this case, $\mathbf{f}_q^{(2)}$ reduces to

$$\mathbf{f}_q^{(2)} = (ge/2mc)\nabla(\mathbf{s}_q \cdot \mathbf{H}), \tag{23e}$$

where the projection of $\mathbf{s}_q$ along $\mathbf{H}$ has a well-defined value, due to the precessional motion. With this simplification, we may proceed to the integration of Eq. (22a) by following a method similar to that suggested in (II), since now the internal variables do not appear explicitly in the equations of motion for the c.m. Hence we assume that there exists a function w such that

$$\mathbf{v}_q = 2D\nabla w - (e/mc)(\mathbf{A} + \epsilon\nabla^2\mathbf{A}), \tag{24}$$

with

$$\epsilon = \tfrac{1}{6}ga_2^2.$$

Since in the present case the c.m. velocity does not depend explicitly on the internal variables, we have

$$\begin{aligned} &\mathfrak{D}_q m\mathbf{v}_q \\ &= m\left[\frac{\partial v_q}{\partial t} + (\mathbf{v}_q \cdot \nabla)\mathbf{v}_q - iD\nabla^2\mathbf{v}_q\right] \\ &= \nabla\left[2mD\frac{\partial w}{\partial t} + \tfrac{1}{2}m\mathbf{v}_q^2 - imD\nabla \cdot \mathbf{v}_q + V + \epsilon\nabla^2 V\right] \\ &\quad + \frac{e}{c}[\mathbf{v}_q - iD\nabla] \times [\mathbf{H} + \epsilon\nabla^2\mathbf{H}] + e\mathbf{E} + e\epsilon\nabla^2\mathbf{E}. \end{aligned}$$

The second equality has been obtained by using Eq. (24), the relation $(e/c)\partial\mathbf{A}/\partial t = -\nabla V - e\mathbf{E}$ and a couple of vectorial identities. If we introduce this result into Eq. (22a) with $\mathbf{f}_q$ given by Eqs. (23), we get the gradient of a function, which yields upon integration

$$2mD\frac{\partial w}{\partial t} + \tfrac{1}{2}m\mathbf{v}_q^2 - imD\nabla \cdot \mathbf{v}_q + V + \epsilon\nabla^2 V - \frac{ge}{2mc}\mathbf{s}_q \cdot \mathbf{H} = 0 \tag{25}$$

if the constant of integration is taken as zero, i.e., if all constants are absorbed in the energy term.[4,13] This differential equation may be linearized by the change of variable

$$w = -i \ln \psi. \tag{26}$$

In fact, rewriting Eq. (25) in terms of ψ with the help of Eqs. (24) and (26), we obtain, after some simple transformations,

$$i\hbar\frac{\partial\psi}{\partial t} = \frac{1}{2m}\left[-i\hbar\nabla - \frac{e}{c}(\mathbf{A} + \epsilon\nabla^2\mathbf{A})\right]^2\psi + (V + \epsilon\nabla^2 V)\psi - \frac{ge}{2mc}\mathbf{s}_q \cdot \mathbf{H}\psi. \tag{27}$$

This is a generalized form of Pauli's equation for a nonrelativistic extended particle with arbitrary spin; here, however, the amplitude ψ is a function of all six coordinates which, according to Eqs. (24) and (26), must be written as a product of a function of the

c.m. coordinates by a function of the internal variables. As stated above, Eq. (27) is written for a given value of $\mathbf{s}_q \cdot \mathbf{H}$; in the general case, however, this product may attain different values. To describe the latter situation we take advantage of the linear character of this equation, which allows us to add up the solutions for different spin states. The amplitude ψ becomes thus a sum of products of functions, a result that will be used in the next section [see Eq. (34)]. It seems convenient to stress at this point that in so doing we are going over from the one-particle problem to that of several particles in different states, employing a procedure analogous to that used to treat the two-body problem in paper IV.

V. SPIN SPECTRUM AND PAULI'S EQUATION

We proceed to demonstrate that the particle's spin is quantized and may take any integral or half-integral value, according to our theory. The proof is actually very simple, since it suffices to show that $\mathbf{s}_q$ may be written as a linear operator which satisfies the properties usually assigned to angular momentum operators; once this is verified, usual angular momentum theory leads us to the desired results.

In (I) we have shown that the usual quantum-mechanical definition of momentum operator follows from the stochastic theory. Hence, we write for particle α

$$\mathbf{p}^\alpha = -i\hbar\nabla_\alpha, \tag{28}$$

and the total momentum operator becomes simply

$$\hat{\mathbf{p}} = \sum_\alpha \hat{\mathbf{p}}^\alpha = -i\hbar\nabla. \tag{29}$$

Similarly, the total angular momentum operator is given by

$$\begin{aligned} \hat{j}_i &= \sum_{\alpha,j,k} \epsilon_{ijk} r_j^\alpha \hat{p}_k^\alpha \\ &= -i\hbar \sum_{j,k} \epsilon_{ijk} \left[r_j \frac{\partial}{\partial r_k} + \sum_\alpha \rho_j^\alpha \frac{\partial}{\partial \rho_k^\alpha} \right]. \end{aligned} \tag{30a}$$

The second equality may be easily demonstrated by making use of the formula

$$\frac{\partial}{\partial r_i^\alpha} = \lambda^\alpha \frac{\partial}{\partial r_i} + \frac{\partial}{\partial \rho_i^\alpha} - \lambda^\alpha \sum_\beta \frac{\partial}{\partial \rho_i^\beta}.$$

To evaluate the last term of Eq. (30a) recall that, owing to the constraints, it represents the contribution of a triad rotating with the body, with its origin at the c.m. Hence, we have[3,6,11]

$$\hat{\mathbf{j}} = -i\hbar \mathbf{r} \times \nabla - i\hbar\nabla_\xi, \tag{30b}$$

where ∇_ξ is a symbol for the operator with components

$$\begin{aligned} (\nabla_\xi)_x &= \frac{\partial}{\partial \alpha_x} \\ &= -\cos\alpha \operatorname{ctg}\beta \partial_\alpha - \sin\alpha \partial_\beta + \cos\alpha \csc\beta \partial_\gamma, \\ (\nabla_\xi)_y &= \frac{\partial}{\partial \alpha_y} \\ &= -\sin\alpha \operatorname{ctg}\beta \partial_\alpha + \cos\alpha \partial_\beta + \sin\alpha \csc\beta \partial_\gamma, \\ (\nabla_\xi)_z &= \frac{\partial}{\partial \alpha_z} = \partial_\alpha. \end{aligned}$$

α_x, α_y, α_z are the angles measured about the fixed x, y, z axes, respectively, and α, β, γ stand for the Euler angles as defined, e.g., in Ref. 11. Therefore, the spin operator is

$$\hat{\mathbf{s}} = -i\hbar\nabla_\xi \tag{31}$$

and its components satisfy the usual commutation relations:

$$[\hat{s}_i, \hat{s}_j] = i\hbar\epsilon_{ijk}\hat{s}_k, \tag{32a}$$

$$[\hat{\mathbf{s}}^2, \hat{s}_k] = 0. \tag{32b}$$

As is well known, we may construct simultaneous eigenfunctions of $\hat{\mathbf{s}}^2$, $\hat{s}_3$, and $\hat{s}_{3'}$, 3′ referring to the body z axis, with eigenvalues $\hbar^2 j(j+1)$, $\hbar m$, and $\hbar k$, respectively, where $j = 0, \frac{1}{2}, 1, \cdots$ and $m, k = -j, -j+1, \cdots, j$ for any given j. Here we want to stress that j may take half-integral values, owing to the fact that the body is constructed from many (> 2) rigidly connected particles and, hence, all three Euler angles are needed to specify the rotations of the system.[13,14] In the case of only one particle (or two rigidly connected point particles with their c.m. at rest), two angles are sufficient to define the most general rotation and, hence, only integral eigenvalues are allowed; this applies, for example, to the orbital angular momentum of the c.m., as is well known.

The angular momentum eigenfunctions are, up to a normalization constant, the generalized spherical or Wigner functions[11] $D_{mk}^{(j)}(\xi)$, where $\xi = (\alpha, \beta, \gamma)$ stand for the three Euler angles. Accordingly, we have

$$\hat{\mathbf{s}}^2 D_{mk}^{(j)}(\xi) = \hbar^2 j(j+1) D_{mk}^{(j)}(\xi), \tag{33a}$$

$$\hat{s}_3 D_{mk}^{(j)}(\xi) = \hbar m D_{mk}^{(j)}(\xi), \tag{33b}$$

$$\hat{s}_\pm D_{mk}^{(j)}(\xi) = \hbar a_{jm}^{(\pm)} D_{m\pm1,k}^{(j)}(\xi), \tag{33c}$$

where

$$\hat{s}_\pm = (2)^{-\frac{1}{2}}(\hat{s}_1 \pm i\hat{s}_2) \tag{33d}$$

and

$$a_{jm}^{(\pm)} = [\tfrac{1}{2}(j \mp m)(j \pm m + 1)]^{\frac{1}{2}}. \tag{33e}$$

Since for a given j any allowed $j_3 = m$ may be realized, the general solution to Eq. (27) for a given k may be

written as follows[1,11]:

$$\psi(\mathbf{r}, \xi) = \sum_{m=-j}^{j} D^{(j)}_{mk}(\xi)\psi_m(\mathbf{r}) \tag{34}$$

(here we are absorbing the unessential normalization constants in the D's). We introduce this $\psi(\mathbf{r}, \xi)$ into Eq. (27), written for simplicity in the form

$$\left(i\hbar\frac{\partial}{\partial t} - H_0\right)\psi = -\frac{ge}{2mc}\,\mathbf{H}\cdot\mathbf{s}_q\psi, \tag{35}$$

and get upon substitution of $\mathbf{s}_q$ by the corresponding operators

$$\sum_m \left\{\left(i\hbar\frac{\partial}{\partial t} - H_0\right)\psi_m(\mathbf{r})\right\} D^{(j)}_{mk}(\xi) = -\frac{ge}{2mc}\sum_{m,i}\{H_i\delta_i D^{(j)}_{mk}(\xi)\psi_m(\mathbf{r})\}. \tag{36}$$

Applying to this equation the formulas Eq. (33) and taking the linear independence of the D's into account, we obtain a system of differential equations for the $\psi_m(\mathbf{r})$, i.e., the c.m. amplitudes for each value of m, thus separating c.m. and interval variables.

Let us write down explicitly the results for the simplest nontrivial case, i.e., for spin $\frac{1}{2}$. Denoting $\psi_{\pm\frac{1}{2}}$ by $\psi_\pm$, we have

$$\psi = D^{(\frac{1}{2})}_{\frac{1}{2}k}(\xi)\psi_+(\mathbf{r}) + D^{(\frac{1}{2})}_{-\frac{1}{2}k}(\xi)\psi_-(\mathbf{r}), \tag{37}$$

with $k = \pm\frac{1}{2}$. Equation (36) then gives us the system

$$\begin{aligned} i\hbar\frac{\partial}{\partial t}\psi_+ &= H_0\psi_+ - \frac{ge\hbar}{4mc}(H_3\psi_+ + H_-\psi_-), \\ i\hbar\frac{\partial}{\partial t}\psi_- &= H_0\psi_- - \frac{ge\hbar}{4mc}(H_+\psi_+ - H_3\psi_-), \end{aligned} \tag{38}$$

where

$$H_\pm = H_1 \pm iH_2, \tag{39}$$

or, in matrix notation,

$$i\hbar\frac{\partial\Psi}{\partial t} = H_0\Psi - \frac{ge\hbar}{4mc}\,\boldsymbol{\sigma}\cdot\mathbf{H}\Psi, \tag{40}$$

σ_i standing for the usual Pauli matrices and Ψ being the two-component amplitude

$$\Psi = \begin{pmatrix}\psi_+\\ \psi_-\end{pmatrix}. \tag{41}$$

In the particular case in which the structure term in H_0 is neglected, i.e., ϵ is taken equal to zero, Eq. (40) reduces itself to Pauli's equation.

VI. SOME COMMENTS ON THE VALIDITY OF THE APPROXIMATION

It is well known[2] that a nonrelativistic model of the type used in this paper is not free from difficulties. For example, if we estimate the minimum value of a_2 needed to guarantee that the energy associated to the spin of value $\frac{1}{2}$ does not exceed the rest energy of the electron, we get $a_2 \gtrsim \hbar/mc$. Similar results may be arrived at by other analogous estimates. This value is very large compared with the classical electron radius $a_0 = \alpha\hbar/mc$, α being the fine structure constant. In other words, even if we consider an electron much more extended than the classical one, we are still in the limits of applicability of nonrelativistic mechanics. However, such large electron radii are not entirely extraneous to modern physics. For example, the fluctuations of the electron's coordinate due to Zitterbewegung are estimated to be precisely of this order of magnitude,[15] i.e., $\langle(\delta x)^2\rangle^{\frac{1}{2}} \sim \hbar/mc$. Since these fluctuations ascribe to the particle an effective radius, we see that both results are in qualitative agreement.

If we use this value of a_2 to estimate the correction to the potential energy due to the finite size of the particle, according to Eq. (27) we get

$$\delta V = \tfrac{1}{6}ga_2^2\nabla^2 V \sim \tfrac{1}{3}(\hbar/mc)^2\nabla^2 V. \tag{42}$$

This result has the same form, sign, and magnitude as the Darwin term which may be deduced from Dirac's equation. Since the Darwin term is interpreted as a consequence of the finite effective radius of the electron,[15] it seems reasonable to consider that Eq. (42) gives a nonrelativistic counterpart of the Darwin term.

VII. RELATIVISTIC GENERALIZATION

In this section we extend the preceding theory to give it a relativistic form by the simple expedient of formally rewriting the equations in explicitly covariant form. As was stated in the introduction, in what concerns the relativistic theory of the stochastic point particle, it seems that this formal procedure simply reproduces the results obtained from a more careful analysis[7]; however, as soon as we try to think of an extended relativistic particle, many difficult questions arise. For example, the very notions of center of mass, rigidity, and a good many related questions must be carefully revised and interpreted to construct a consistent theory. At the time being there are several interesting and ingenious treatments of the problem; however, some of these theories are not only not equivalent, but even conflicting among themselves. Since it is not the purpose of this paper to go more deeply into such questions, we here content ourselves with a first attempt by formally writing down the nonrelativistic spin terms in their corresponding explicitly covariant form. The procedure, although not entirely satisfactory, allows us to get some interesting results which would be otherwise much more

difficult to attain and may perhaps be more questionable.

To begin with, we introduce the (adimensional) 4-velocities v_μ and u_μ as relativistic extensions of $\mathbf{v}$ and $\mathbf{u}$; from them we construct the systematic (4) momentum p_μ and the stochastic (4) momentum q_μ as follows:

$$p_\mu = mcv_\mu,$$
$$q_\mu = mcu_\mu. \tag{43}$$

In terms of these momenta, Eqs. (3) generalize to give

$$m\mathfrak{D}_c = p_\mu\partial_\mu,$$
$$m\mathfrak{D}_s = q_\mu\partial_\mu + mD\partial_\mu\partial_\mu, \tag{44}$$

where the sum convention over repeated Greek indices is being used and μ takes the values 1 to 4 with $x_\mu = (\mathbf{x}, ix_0)$, $x_0 = ct$. Then Eqs. (14) become

$$\mathfrak{D}_c p_\mu - \mathfrak{D}_s q_\mu = f^{(+)}_{0\mu},$$
$$\mathfrak{D}_s p_\mu + \mathfrak{D}_c q_\mu = f^{(-)}_{0\mu}. \tag{45}$$

The 4-forces $f^{(\pm)}_{0\mu}$ may be calculated for a point particle with methods similar to those used in the nonrelativistic case. To all orders, the results for the electromagnetic case are the covariant generalization of Eqs. (16′)[7]:

$$f^{(+)}_{0\mu} = \frac{e}{mc}F_{\mu\nu}p_\nu - \frac{e}{c}\mathfrak{D}_c A_\mu + \frac{e}{mc}p_\nu\partial_\nu A_\mu,$$
$$f^{(-)}_{0\mu} = \frac{e}{mc}F_{\mu\nu}q_\nu - \frac{e}{c}\mathfrak{D}_s A_\mu + \frac{e}{mc}q_\nu\partial_\nu A_\mu, \tag{46}$$

where $F_{\mu\nu}$ stands for the electromagnetic tensor.

Since the operators $\mathfrak{D}_c$ and $\mathfrak{D}_s$ have been written above only to second order, we write Eqs. (46) to the same order, thus getting

$$f^{(+)}_{0\mu} = (e/mc)F_{\mu\nu}p_\nu,$$
$$f^{(-)}_{0\mu} = (e/mc)F_{\mu\nu}q_\nu + (eD/c)\partial_\nu F_{\mu\nu}. \tag{47}$$

These results represent the covariant generalization of Eq. (16), written for the gauge $\partial_\mu A_\mu = 0$. Hence, taking into account the spin of the particle, but neglecting terms analogous to $\mathbf{f}^{(3)}_q$ [cf. Eq. (23d)], we write as our fundamental system of relativistic equations the following set:

$$\mathfrak{D}_c p_\mu - \mathfrak{D}_s q_\mu = (e/mc)F_{\mu\nu}p_\nu + (ge/4mc)\partial_\mu s^c_{\lambda\nu}F_{\lambda\nu},$$
$$\mathfrak{D}_c q_\mu + \mathfrak{D}_s p_\mu = (e/mc)F_{\mu\nu}q_\nu + (eD/c)\partial_\nu F_{\mu\nu} + (ge/4mc)\partial_\mu s^s_{\lambda\nu}F_{\lambda\nu}. \tag{48}$$

By introducing the complex variables

$$\mathfrak{P}_\mu = p_\mu - iq_\mu,$$
$$S^q_{\mu\nu} = s^c_{\mu\nu} - is^s_{\mu\nu}, \tag{49}$$
$$f^q_\mu = f^{(+)}_{0\mu} - if^{(-)}_{0\mu} + (ge/4mc)\partial_\mu S^q_{\lambda\nu}F_{\lambda\nu},$$

we can rewrite Eqs. (48) in the form

$$\mathfrak{D}_q\mathfrak{P}_\mu = f^q_\mu \tag{50a}$$

and its c.c. Explicitly,

$$f^q_\mu = (e/mc)F_{\mu\nu}\mathfrak{P}_\nu - (iDe/c)\partial_\nu F_{\mu\nu} + (ge/4mc)\partial_\mu S^q_{\lambda\nu}F_{\lambda\nu}. \tag{50b}$$

To integrate Eqs. (50), first note the relation

$$[m\mathfrak{D}_q, \partial_\lambda] = -(\partial_\lambda\mathfrak{P}_\mu)\partial_\mu, \tag{51}$$

and then write $\mathfrak{P}_\mu$ in the form

$$\mathfrak{P}_\mu = \hbar\partial_\mu w - (e/c)A_\mu, \tag{52}$$

assuming that such a w exists. Applying $\mathfrak{D}_q$ to Eq. (52), using Eq. (51) to calculate $\partial_\lambda\mathfrak{D}_q w$, and introducing into this result Eq. (59b), we get

$$\mathfrak{D}_q\mathfrak{P}_\mu = \hbar\partial_\mu\mathfrak{D}_q w - (\hbar^2/2m)\partial_\mu(\partial_\nu w\partial_\nu w) + f^q_\mu + (e^2/2mc^2)\partial_\mu(A_\nu A_\nu) - (ge/4mc)\partial_\mu S^q_{\lambda\nu}F_{\lambda\nu}.$$

Therefore, from (50a) we obtain, after an elementary integration:

$$2\hbar m\mathfrak{D}_q w - \hbar^2(\partial_\mu w)(\partial_\mu w) + \frac{e^2}{c^2}A_\mu A_\mu = -m^2c^2 + (ge/2c)S^q_{\mu\nu}F_{\mu\nu}. \tag{53}$$

The constant of integration has been selected so that in the limit of a spinless nonstochastic particle we obtain from Eq. (53) the relativistic relation $p_\mu p_\mu = -m^2c^2$. If now we write this result in terms of the amplitude ψ previously defined, Eq. (26), we get

$$\left(-i\hbar\partial_\mu - \frac{e}{c}A_\mu\right)^2\psi + m^2c^2\psi = (ge/2c)S^q_{\mu\nu}F_{\mu\nu}\psi. \tag{54}$$

This expression has the form of the Feynman–Gell-Mann equation,[9] usually considered valid for any spin, but with the spin written still in terms of the relativistic Euler angles.[5,6,8]

A further rigorous treatment of the problem along lines similar to that followed in the nonrelativistic case requires explicit use of the theory of the relativistic spinning rigid body. To simplify matters, we prefer making an approximation which allows us to use the nonrelativistic results. Let us consider the case in which the electric field may be neglected, i.e., let us take $A_0 = 0$ and, hence, $\partial_t\mathbf{A} = 0$ and write

$$S^q_{\mu\nu}F_{\mu\nu} \approx 2\mathbf{S}^q \cdot \mathbf{H}. \tag{55}$$

Introducing this approximation into Eq. (54) and proceeding as in Sec. V, we get for $s = \frac{1}{2}$

$$\left(-i\hbar\partial_\mu - \frac{e}{c}A_\mu\right)^2\Phi + m^2c^2\Phi = (ge\hbar/2c)\boldsymbol{\sigma}\cdot\mathbf{H}\Phi, \tag{56a}$$

with

$$\Phi = \begin{pmatrix}\psi_+\\ \psi_-\end{pmatrix}. \tag{56b}$$

Known algebraic transformations[15] may be used to cast this second-order equation for a two-component amplitude into a first-order equation for a four-component amplitude. In fact, defining φ and χ by

$$\Phi = \tfrac{1}{2}(\varphi - \chi), \tag{57a}$$

$$\left[i\hbar\partial_0 - \boldsymbol{\sigma}\cdot\left(i\hbar\nabla + \frac{e}{c}\mathbf{A}\right)\right]\Phi = \tfrac{1}{2}mc(\varphi + \chi), \tag{57b}$$

factorization of Eq. (56a) gives

$$-i\hbar\partial_0\varphi - \boldsymbol{\sigma}\cdot\left(i\hbar\nabla + \frac{e}{c}\mathbf{A}\right)\chi = -mc\varphi + \frac{\hbar e}{4mc^2}(g-2)\boldsymbol{\sigma}\cdot\mathbf{H}(\varphi - \chi)$$

and

$$\boldsymbol{\sigma}\cdot\left(i\hbar\nabla + \frac{e}{c}\mathbf{A}\right)\varphi + i\hbar\partial_0\chi = -mc\chi + \frac{\hbar e}{4mc^2}(g-2)\boldsymbol{\sigma}\cdot\mathbf{H}(\chi - \varphi).$$

Hence, if we introduce the matrices

$$\gamma_k = \begin{pmatrix} 0 & -i\sigma_k \\ i\sigma_k & 0 \end{pmatrix}, \quad \gamma_4 = \begin{pmatrix} I & 0 \\ 0 & -I \end{pmatrix}, \tag{58a}$$

$$\Sigma_k = \begin{pmatrix} \sigma_k & 0 \\ 0 & \sigma_k \end{pmatrix}, \qquad \gamma_5 = -\begin{pmatrix} 0 & I \\ I & 0 \end{pmatrix}, \tag{58b}$$

and the four-component amplitude

$$\Psi = \begin{pmatrix} \varphi \\ \chi \end{pmatrix}, \tag{59}$$

Eq. (56a) may be written in the form

$$i\gamma_\mu\left(-i\hbar\partial_\mu - \frac{e}{c}A_\mu\right)\Psi + mc\Psi = \frac{\hbar e}{4mc^2}(g-2)(1+\gamma_5)\boldsymbol{\Sigma}\cdot\mathbf{H}\Psi. \tag{60}$$

We see that if $g \neq 2$, a term proportional to $1 + \gamma_5$ appears in the final result; hence, from conservation of parity it follows that we must take $g = 2$. We see that the proposed relativistic generalization of the stochastic theory predicts a correct value for the gyromagnetic ratio of the electron and that it leads directly to Dirac's equation for the spin $\frac{1}{2}$ case. When we take $g = 2$, Eq. (60) recovers a fully covariant form and so the restriction over the electric field may be removed.

VIII. CONCLUSIONS

In the series of papers to which the present one belongs, we have attempted to demonstrate that a stochastic theory related to a subquantum level may be used as the basis for the foundations of usual quantum mechanics, and that this approach has the advantages of simplicity, physical clarity, and economy of postulates. However, emphasis must be given to the fact that our theory differs considerably from usual quantum mechanics in its conceptual framework and, hence, that the physical content of many of the results is not the same in both theories; concrete examples are the interpretation of the uncertainty relationships [see (I)] or, still, some aspects of the two-particle problem [see (IV)]. In the present paper we have a third example in which the spin is treated in a nonorthodox manner from the standpoint of usual quantum mechanics. However, the theory is still incomplete and many important questions remain open; among them, we may mention the very fundamental problem of constructing the theory of the subquantal, stochastic interaction, the lack of which implies the need to use phenomenological parameters. The possibility of extending the theory so as to cover quantum electrodynamics seems realistic; moreover, some work along this direction may be found in the literature.[16]

[1] L. de la Peña-Auerbach, Phys. Letters **27A**, 594 (1968), and J. Math. Phys. **10**, 1620 (1969) [to be referred to as (I)]; L. de la Peña-Auerbach and A. M. Cetto, Rev. Mex. Fís. **17**, 327 (1968) [to be referred to as (II)]; Phys. Letters **29A**, 562 (1969), and Rev. Mex. Fís. **18**, 253 (1969) [to be referred to as (III)]; Rev. Mex. Fís. **18**, 323 (1969) [to be referred to as (IV)].

[2] M. Jammer, *The Conceptual Development of Quantum Mechanics* (McGraw-Hill, New York, 1966).

[3] N. Rosen, Phys. Rev. **82**, 621 (1951).

[4] A. Sparzani, Nuovo Cimento Suppl. **4**, 231 (1966).

[5] D. Bohm, P. Hillion, T. Takabayasi, and J. P. Vigier, Progr. Theoret. Phys. **23**, 496 (1960).

[6] O. Hara, and T. Gotō, Suppl. Progr. Theoret. Phys. (Kyoto) **41**, 56 (1968).

[7] L. de la Peña-Auerbach, Rev. Mex. Fís. **19**, 133 (1970).

[8] For a comparative study of several theories, see P. Nyborg, Nuovo Cimento **23**, 47 (1962), and **26**, 821 (1962). For a representative list of the literature on the subject, see, e.g., H. C. Corben, *Classical and Quantum Theories of Spinning Particles* (Holden-Day, Inc., San Francisco, 1968).

[9] R. P. Feynman and R. Gell-Mann, Phys. Rev. **109**, 193 (1958).

[10] We have restricted our study to the quantum-mechanical case by writing Eqs. (1) and (3) in the corresponding particular form. For details see (I).

[11] See, e.g., A. P. Edmonds, *Angular Momentum in Quantum Mechanics* (Princeton U.P., Princeton, N.J., 1957).

[12] These definitions differ from those used in (II) and (III) by a factor $-i$; hence, we are using a lower case subscript q instead of Q to avoid confusion.

[13] F. Bopp and R. Haag, Z. Naturforsch. **5a**, 644 (1950).

[14] W. Pauli, "Die Allgemeinen Prinzipien der Wellenmechanik," *Handbuch der Physik, Band 24, Teil 1* (Springer-Verlag, Berlin, 1933).

[15] See, e.g., J. J. Sakurai, *Advanced Quantum Mechanics* (Addison-Wesley, New York, 1967).

[16] A point of view close to ours is adopted by T. H. Boyer, Phys. Rev. **174**, 1631 (1968), based on a stochastic theory proposed by D. Kershaw, Phys. Rev. **136**, B1850 (1964). A discussion of Kershaw's theory may be found in L. de la Peña-Auerbach and R. M. Velasco, Rev. Mex. Fís. **18**, 397 (1969).

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 12, NUMBER 3 MARCH 1971

Nonrelativistic Current Algebras as Unitary Representations of Groups*

GERALD A. GOLDIN
Department of Physics, University of Pennsylvania, Philadelphia, Pennsylvania 19104

(Received 12 August 1970)

It is possible that a complete physical theory can be written entirely in terms of operators such as current densities, rather than in terms of field operators. The current densities in these models correspond in general to distributions of unbounded operators. Such a theory is reviewed for the case of nonrelativistic quantum mechanics. It is found that one can exponentiate the current algebra to obtain a group, which may then be represented by unitary (hence bounded) operators in Hilbert space. This procedure is analogous to exponentiating the canonical commutation relations and obtaining the Weyl group. For nonrelativistic quantum mechanics without spin, the group is the semidirect product $\mathfrak{J} \wedge \mathfrak{K}$ Schwartz space $\mathfrak{J}(\mathbb{R}^3)$ with a group $\mathfrak{K}$ of certain C_∞ diffeomorphisms from $\mathbb{R}^3$ onto itself. For $f \in \mathfrak{J}$ and $\varphi \in \mathfrak{K}$, the composition mapping $(f, \varphi) \to f \circ \varphi$ defines the semidirect product law. The Gel'fand–Vilenkin formalism for "nuclear Lie groups" is suitable for the representation theory of such a group. Almost all of the physical information is contained in a cylindrical measure μ on $\mathfrak{J}'$, the dual of the nuclear space $\mathfrak{J}$. The Fourier transform of μ can be interpreted as an expectation functional with respect to the state of lowest energy. The nonrelativistic Fock representation (including the theory of n particles in a box) is examined in this formalism. Conditions on μ are systematically developed which suffice to recover all of the infinitesimal generators on a common, dense, invariant domain in the Hilbert space. For the Weyl group slightly weaker conditions than for the semidirect product case would suffice. Gaussian measures in $\mathfrak{J}'$, as well as the measures defining n-particle representations of $\mathfrak{J} \wedge \mathfrak{K}$, satisfy these conditions.

1. INTRODUCTION

In a recent series of papers, Dashen, Sharp, and Callan investigate the possibility that a complete physical theory can be written entirely in terms of operators such as current densities, rather than in terms of field operators.[1–3] They approach this question through the study of various models. Typically, the current densities are defined in terms of field operators, and their commutation relations are computed under the assumption that the fields satisfy equal-time canonical commutation or anticommutation relations. But once the equal-time current commutators have been obtained, these are taken as the new starting point towards a theory of the physical world.[4]

The main purpose of the present paper, which is based on the author's doctoral thesis, is to develop a rigorous mathematical foundation for the nonrelativistic current algebra.[5]

Like the more familiar field operators, the current densities will correspond (in general) to distributions of unbounded operators in a Hilbert space $\mathcal{H}$. The consideration of such a representation involves us in familiar mathematical difficulties. The smeared current may have a domain of definition, dependent on the testing function, which is not all of $\mathcal{H}$; thus the commutator of two currents, for example, may not even be defined.

These difficulties are overcome in the present approach by looking for representations of the *exponentiated* commutation relations. The *groups* thus obtained from the equal-time current algebras are represented by unitary (hence bounded) operators in $\mathcal{H}$. This procedure is analogous to exponentiating the canonical commutation relations and obtaining the Weyl group.[6]

Let us focus our discussion by presenting the nonrelativistic model under consideration. Using the equal-time canonical commutation or anticommutation relations of nonrelativistic field theory, one first proceeds formally in order to obtain the current algebra. The computations are reformulated rigorously in the Fock representation in Sec. 2.

A. Model for Nonrelativistic Quantum Mechanics[1]

Suppose that the second-quantized fields $\psi(\mathbf{x})$ and $\psi^*(\mathbf{x})$ at a fixed time t satisfy the nonrelativistic equal-time canonical commutation relations

$$[\psi(\mathbf{x}), \psi(\mathbf{y})]_- = [\psi^*(\mathbf{x}), \psi^*(\mathbf{y})]_- = 0,$$
$$[\psi(\mathbf{x}), \psi^*(\mathbf{y})]_- = \delta(\mathbf{x} - \mathbf{y}) \tag{1.1}$$

or the anticommutation relations[7]

$$[\psi(\mathbf{x}), \psi(\mathbf{y})]_+ = [\psi^*(\mathbf{x}), \psi^*(\mathbf{y})]_+ = 0,$$
$$[\psi(\mathbf{x}), \psi^*(\mathbf{y})]_+ = \delta(\mathbf{x} - \mathbf{y}). \tag{1.2}$$

We define the "particle density"

$$\rho(\mathbf{x}) = \psi^*(\mathbf{x})\psi(\mathbf{x}) \tag{1.3}$$

and the "current density"

$$\mathbf{J}(\mathbf{x}) = (2i)^{-1}\{\psi^*(\mathbf{x})\nabla\psi(\mathbf{x}) - [\nabla\psi^*(\mathbf{x})]\psi(\mathbf{x})\}. \tag{1.4}$$

Using the relations (1.1) or (1.2), we can now compute the commutation relations satisfied by ρ and $\mathbf{J}$. These turn out to be

$$[\rho(\mathbf{x}), \rho(\mathbf{y})] = 0, \tag{1.5}$$

$$[\rho(\mathbf{x}), J_k(\mathbf{y})] = -i\frac{\partial}{\partial x^k}[\delta(\mathbf{x}-\mathbf{y})\rho(\mathbf{x})], \tag{1.6}$$

and

$$[J_j(\mathbf{x}), J_k(\mathbf{y})] = -i\frac{\partial}{\partial x^k}[\delta(\mathbf{x}-\mathbf{y})J_j(\mathbf{x})] + i\frac{\partial}{\partial y^j}[\delta(\mathbf{x}-\mathbf{y})J_k(\mathbf{y})], \tag{1.7}$$

independently of whether one starts with (1.1) or (1.2).

Let us obtain the smeared form of Eqs. (1.5)–(1.7).[8] For notational purposes, let $\mathbf{g}$ denote a triple (g_1, g_2, g_3) of smearing functions, and define

$$J(\mathbf{g}) = \sum_{k=1}^{3} J_k(g_k), \tag{1.8}$$

where as usual

$$J_k(f) = \int J_k(\mathbf{x}) f(\mathbf{x})\, d\mathbf{x}, \tag{1.9}$$

etc. The smeared commutation relations then become

$$[\rho(f), \rho(g)] = 0, \tag{1.10}$$

$$[\rho(f), J(\mathbf{g})] = i\rho(\mathbf{g}\cdot\nabla f), \tag{1.11}$$

$$[J(\mathbf{f}), J(\mathbf{g})] = iJ(\mathbf{g}\cdot\nabla\mathbf{f} - \mathbf{f}\cdot\nabla\mathbf{g}). \tag{1.12}$$

It is also interesting to consider the nonrelativistic model with spin. Suppose now that the fields $\psi(\mathbf{x})$ and $\psi^*(\mathbf{x})$ are two-component spinors

$$\psi = \begin{bmatrix}\psi_1\\ \psi_2\end{bmatrix}, \quad \psi^* = [\psi_1^*; \psi_2^*] \tag{1.13}$$

satisfying either canonical commutation (−) or anticommutation (+) relations

$$[\psi_r(\mathbf{x}), \psi_s(\mathbf{y})]_\pm = [\psi_r^*(\mathbf{x}), \psi_s^*(\mathbf{y})]_\pm = 0,$$
$$[\psi_r(\mathbf{x}), \psi_s^*(\mathbf{y})]_\pm = \delta_{rs}\delta(\mathbf{x}-\mathbf{y}). \tag{1.14}$$

Then, in addition to the "particle density" $\rho(\mathbf{x})$ and the "current density" $\mathbf{J}(\mathbf{x})$ defined by (1.3) and (1.4), a "spin density"

$$\mathbf{\Sigma}(\mathbf{x}) = \tfrac{1}{2}\psi^*(\mathbf{x})\boldsymbol{\sigma}\psi(\mathbf{x}) \tag{1.15}$$

must be introduced, where $\boldsymbol{\sigma} = (\sigma_1, \sigma_2, \sigma_3)$ are the familiar Pauli spin matrices. These, we recall, satisfy the commutation relations

$$[\sigma_j, \sigma_k] = 2i\varepsilon_{jkl}\sigma_l. \tag{1.16}$$

The distributions ρ and J continue to satisfy the algebra given by Eqs. (1.5)–(1.7) or (1.10)–(1.12). In addition, we have the new commutation relations

$$[\rho(\mathbf{x}), \Sigma_j(\mathbf{y})] = 0, \tag{1.17}$$

$$[\Sigma_j(\mathbf{x}), J_k(\mathbf{y})] = -i\frac{\partial}{\partial x^k}[\delta(\mathbf{x}-\mathbf{y})\Sigma_j(\mathbf{x})], \tag{1.18}$$

$$[\Sigma_j(\mathbf{x}), \Sigma_k(\mathbf{y})] = i\varepsilon_{jkl}\delta(\mathbf{x}-\mathbf{y})\Sigma_l(\mathbf{y}), \tag{1.19}$$

again independently of whether we start with commutation or anticommutation relations in (1.4). Letting

$$\Sigma(\mathbf{g}) = \sum_{k=1}^{3}(\Sigma_k(g_k)), \tag{1.20}$$

and recalling that

$$\sum_{j,k=1}^{3}\varepsilon_{jkl}f_j g_k = (\mathbf{f}\times\mathbf{g})_l, \tag{1.21}$$

we find that the smeared commutation relations become

$$[\rho(f), \Sigma(\mathbf{g})] = 0, \tag{1.22}$$

$$[\Sigma(\mathbf{f}), J(\mathbf{g})] = i\Sigma(\mathbf{g}\cdot\nabla\mathbf{f}), \tag{1.23}$$

$$[\Sigma(\mathbf{f}), \Sigma(\mathbf{g})] = i\Sigma(\mathbf{f}\times\mathbf{g}). \tag{1.24}$$

B. Relativistic Models[1–3]

The status of the various relativistic models remains open. In the case of nonrelativistic quantum mechanics, the currents can be obtained rigorously from fields in the Fock representation; but in the relativistic models, no such representations of the current algebras are immediately available.

Furthermore, it is well known that Schwinger terms cannot be neglected in relativistic models.[5,9] But the problem of representing an equal-time algebra in which the current commutators may be formally infinite (such as occurs in the quark model) remains to be overcome.

At present these are the major considerations in attempting to generalize from the nonrelativistic case discussed in the present paper.

C. Discussion

Now that we have considered, in the nonrelativistic model, the local operators which are taken to be the "coordinates" or "building blocks" of a physical theory, let us ask what is necessary to specify such a theory completely.

First, we need a representation of the (smeared) current commutation relations by self-adjoint operators in a Hilbert space $\mathcal{H}$. As we shall see in Sec. 3, we may require instead a *unitary* representation of the *group* obtained by exponentiating the current commutators.

Next, we expect there to be a unitary representation E in $\mathcal{H}$ of the Euclidean group. This representation must map the fixed-time current algebra into itself via the formula

$$E(\mathbf{a}, R)J(f)E(\mathbf{a}, R)^* = J(f_{(\mathbf{a},R)}) \tag{1.25}$$

for a local current J, where

$$f_{(\mathbf{a},R)}(\mathbf{x}) = f(R^{-1}(\mathbf{x} - \mathbf{a})). \tag{1.26}$$

In addition, there will be a self-adjoint Hamiltonian H which generates time translations; i.e.,

$$[H, \mathcal{O}] = i^{-1}\dot{\mathcal{O}} \tag{1.27}$$

for a fixed-time operator $\mathcal{O}$ built from the currents. The spectrum of H must be bounded below.

Since currents, unlike fields, are in principle directly observable, operators such as the total charge or the baryon number necessarily commute with all of the currents. These operators define superselection rules, and reduce the Hilbert space into mutually incoherent subspaces of fixed charge, baryon number, etc. Suppose that for a certain model, the currents are defined from the fields in a particular representation, and we wish to know when "enough" currents have been defined to give us a "complete set of coordinates." The answer is obtained by finding the superselecting operators in the representation. When the only superselecting operators are those expected from experimental observations, we have included enough observables in the current algebra.

In each irreducible subspace, one may assume the existence of a vector Ω cyclic for some maximal Abelian subset of the currents. For certain potentials in nonrelativistic quantum mechanics, there exists a (normalizable) state of lowest energy Ω. In such theories, the Hamiltonian can be recovered on a dense set from the choice of cyclic vector Ω, $H\Omega = 0$, together with the equation of continuity

$$\frac{\partial \rho}{\partial t} = -\nabla \cdot \mathbf{J}. \tag{1.28}$$

Essentially, all of the physical information is contained in the expectation functional $(\Omega, e^{i\rho(f)}\Omega)$.

For relativistic models, we shall expect that one can in principle perform any physical experiment in the vacuum sector by appropriately introducing particles or antiparticles "behind the moon."[10] Taking the cyclic vector to be the vacuum in the irreducible subspace containing it, we should be able to use a conservation equation to conclude that the matrix elements of the Hamiltonian are determined by a vacuum expectation functional. Similarly, in the vacuum sector of $\mathcal{H}$, one should be able to recover the representation of the Lorentz group from the vacuum expectation functional.[6]

A final remark: One expects that there may be representations of the current algebra which do not arise from an underlying representation of the canonical formalism. If such a representation describes the physical world, then local currents must replace fields as the "building blocks" of nature. In the relativistic axiomatic theory of local observables, one associates with each bounded region of space–time a C^*-algebra of local observables.[10] Our view is that the exponentiated currents and their linear combinations may constitute a promising specific choice for the generators of such a system of C^*-algebras at a fixed time t. One recovers the 4-dimensionally smeared algebras by representing the fixed-time formalism in a Hilbert space, specifying the Hamiltonian, and computing (for a bounded local operator A)

$$A(h) = \int dt e^{iHt}A(h_t)e^{-iHt}, \tag{1.29}$$

where h is a 4-dimensional smearing function, $h_t(\mathbf{x}) = h(\mathbf{x}, t)$, and

$$A(h_t) = \int d\mathbf{x} A(\mathbf{x})h_t(\mathbf{x}). \tag{1.30}$$

The (generally unbounded) currents are *affiliated* in $\mathcal{H}$ with the weak closures of the local C^*-algebras, in a sense which can be made mathematically precise.[11]

2. THE FOCK REPRESENTATION

In this section, we look at the currents defined in the nonrelativistic Fock representation of the fields.[12,13] Thus, one obtains a (reducible) representation of the current algebra [(1.10)–(1.12)], which splits into the direct sum of irreducible representations having fixed particle-number and exchange symmetry. It is demonstrated that in one spatial dimension the n-particle representations of (1.10)–(1.12) having different exchange symmetries are unitarily equivalent, although they are inequivalent in higher dimensions. The Fock representation provides an important example for study; it is useful in exponentiating the current algebra (Sec. 3) and in evaluating the status of the representation theory (Sec. 4).

Let $\mathcal{H}$ denote the Hilbert space. $\Psi \in \mathcal{H}$ may be written $(\Psi_1, \Psi_2, \cdots, \Psi_n, \cdots)$, where

$$\Psi_n = \Psi_n(\mathbf{x}_1, \cdots, \mathbf{x}_n)$$

is an L^2 function of n vector variables symmetric (or antisymmetric, in the case of the anticommutation relations) with respect to coordinate exchange. We

impose the condition

$$(\Psi, \Psi) = \sum_{n=0}^{\infty} (\Psi_n, \Psi_n) < \infty, \tag{2.1}$$

where (Ψ_n, Ψ_n) denotes the usual L^2 inner product.

The action of fields satisfying the commutation relations (1.1) is given by

$$[\psi(f)\Psi]_n(\mathbf{x}_1, \cdots, \mathbf{x}_n) = (n+1)^{\frac{1}{2}} \int \Psi_{n+1}(\mathbf{x}_1, \cdots, \mathbf{x}_n, \mathbf{x}) f(\mathbf{x})\, d\mathbf{x}, \tag{2.2}$$

$$[\psi^*(f)\Psi]_n(\mathbf{x}_1, \cdots, \mathbf{x}_n) = n^{-\frac{1}{2}} \sum_{j=1}^{n} \overline{f(\mathbf{x}_j)} \Psi_{n-1}(\mathbf{x}_1, \cdots, \hat{\mathbf{x}}_j, \cdots, \mathbf{x}_n) \tag{2.3}$$

or, in unsmeared form,

$$[\psi(\mathbf{x})\Psi]_n(\mathbf{x}_1, \cdots, \mathbf{x}_n) = (n+1)^{\frac{1}{2}} \Psi(\mathbf{x}_1, \cdots, \mathbf{x}_n, \mathbf{x}), \tag{2.4}$$

$$[\psi^*(\mathbf{x})\Psi]_n(\mathbf{x}_1, \cdots, \mathbf{x}_n) = n^{-\frac{1}{2}} \sum_{j=1}^{n} \delta(\mathbf{x} - \mathbf{x}_j) \Psi_{n-1}(\mathbf{x}_1, \cdots, \hat{\mathbf{x}}_j, \cdots, \mathbf{x}_n). \tag{2.5}$$

The action of fields satisfying the anticommutation relations (1.2) is

$$[\psi(f)\Psi]_n(\mathbf{x}_1, \cdots, \mathbf{x}_n) = (n+1)^{\frac{1}{2}} \int \Psi_{n+1}(\mathbf{x}_1, \cdots, \mathbf{x}_n, \mathbf{x}) f(\mathbf{x})\, d\mathbf{x}, \tag{2.6}$$

$$[\psi^*(f)\Psi]_n(\mathbf{x}_1, \cdots, \mathbf{x}_n) = \frac{(-1)^{n+l}}{n^{\frac{1}{2}}} \sum_{j=1}^{n} (-1)^{j+1} \overline{f(\mathbf{x}_j)} \Psi_{n-1}(\mathbf{x}_1, \cdots, \hat{\mathbf{x}}_j, \cdots, \mathbf{x}_n), \tag{2.7}$$

or

$$[\psi(\mathbf{x})\Psi]_n(\mathbf{x}_1, \cdots, \mathbf{x}_n) = (n+1)^{\frac{1}{2}} \Psi_{n+1}(\mathbf{x}_1, \cdots, \mathbf{x}_n, \mathbf{x}), \tag{2.8}$$

$$[\psi^*(\mathbf{x})\Psi]_n(\mathbf{x}_1, \cdots, \mathbf{x}_n) = \frac{(-1)^{n+1}}{n^{\frac{1}{2}}} \sum_{j=1}^{n} (-1)^{j+1} \delta(\mathbf{x} - \mathbf{x}_j) \times \Psi_{n-1}(\mathbf{x}_1, \cdots, \hat{\mathbf{x}}_j, \cdots, \mathbf{x}_n). \tag{2.9}$$

Defining ρ and J by (1.3) and (1.4), one obtains, either from (2.4) and (2.5) or from (2.8) and (2.9), the result

$$[\rho(f)\Psi]_n = \sum_{j=1}^{n} f(\mathbf{x}_j) \Psi_n, \tag{2.10}$$

$$[J(\mathbf{g})\Psi]_n = -\tfrac{1}{2} i \sum_{j=1}^{n} [\mathbf{g}(\mathbf{x}_j) \cdot \nabla_j + \nabla_j \cdot \mathbf{g}(\mathbf{x}_j)] \Psi_n \tag{2.11}$$

for the smeared operators. One can verify directly that Eqs. (2.10) and (2.11) indeed define a representation of the current algebra (1.10)–(1.12).

The n-particle subspace of the Fock representation is defined to be the set $\mathcal{H}_n = \{\Psi \in \mathcal{H} \mid \Psi = (0, \cdots, 0, \Psi_n, 0, \cdots)\}$. It is clear from (2.10) and (2.11) that $\rho(f)$ and $J(\mathbf{g})$ map $\mathcal{H}_n$ into $\mathcal{H}_n$; $\rho(f)$ and $J(\mathbf{g})$ commute with the number operator $(N\Psi)_n = n\Psi_n$. The "n-particle representation" means the representation (2.10) and (2.11) restricted to $\mathcal{H}_n$. For $n \geq 2$, we here distinguish between n-particle representations for which the Ψ_n have different exchange symmetries. (We note that ρ and J always preserve the exchange symmetry of Ψ_n.)

In the n-particle representation, $\rho(f)$ is a bounded operator, and $\|\rho(f)\| = n \cdot \sup f(\mathbf{x})$. The $J(\mathbf{g})$, on the other hand, are unbounded operators in $\mathcal{H}_n$. But the dense subspace of $\mathcal{H}_n$ consisting of only the Schwartz-space functions provides a common domain for all of the $\rho(f)$ and $J(\mathbf{g})$; furthermore, this domain is invariant under the action of the ρ's and the J's.

Next we proceed to show the unitary equivalence of n-particle representations of the current algebra (1.10)–(1.12) having different exchange symmetries, in the case of one spatial dimension.

Lemma 1: Let A be a differential operator (defined on $\mathfrak{I}$) which is essentially self-adjoint in $L^2(\mathbb{R}^n)$. A^{**} is its self-adjoint extension. Let $\Psi \in L^2(\mathbb{R}^n)$, and consider $\Psi(x_1, \cdots, x_n)$ as a generalized function.

Then $\Psi \in D_{A^{**}}$ (the domain of A^{**}) if and only if the generalized function $(A\Psi)(x_1, \cdots, x_n)$ is square integrable.

Proof: Suppose $\Psi \in L^2(\mathbb{R}^n)$ with $(A\Psi)(x_1, \cdots, x_n) \in L^2(\mathbb{R}^n)$. Then, for $\Phi \in \mathfrak{I}$, $(\Psi, A\Phi) = (\chi, \Phi)$, where $\chi(x_1, \cdots, x_n) = (A\Psi)(x_1, \cdots, x_n) \in L^2(\mathbb{R}^n)$. Thus $\Psi \in D_{A^*} = D_{A^{**}}$.

The converse is trivial. QED

Lemma 1 can be extended without difficulty to the case where A is a differential operator preserving exchange symmetry and acting on the symmetric or antisymmetric subspace of $L^2(\mathbb{R}^n)$.

Lemma 2: Let $\mathcal{H}_s(\mathcal{H}_a)$ denote the Hilbert space of symmetric (antisymmetric) L^2 functions of n variables. Let $\sigma(x)$ be the generalized function $x/|x|$, and define $\sigma(x_1, \cdots, x_n) = \prod_{r_1 < r_2} \sigma(x_{r_2} - x_{r_1})$ for $n \geq 2$. $\sigma(x_1, \cdots, x_n)$ is totally antisymmetric with respect to coordinate exchange.

Let $J(g)$ be the differential operator

$$\sum_{k=1}^{n} \frac{1}{2i} \left(g(x_k) \frac{\partial}{\partial x_k} + \frac{\partial}{\partial x_k} g(x_k) \right), \tag{2.12}$$

where g has, say, compact support. We shall see in Sec. 3 that $J(g)$ defines an essentially self-adjoint operator on the domain of Schwartz-space functions in L^2, $\mathcal{H}_s$, or $\mathcal{H}_a$.

Then (as generalized functions)

$$J(g)\sigma(x_{r_2} - x_{r_1}) = (2i)^{-1}\sum_{k=1}^{n} g'(x_k)\sigma(x_{r_2} - x_{r_1}), \tag{2.13}$$

$$J(g)\sigma(x_1, \cdots, x_n) = (2i)^{-1}\sum_{k=1}^{n} g'(x_k)\sigma(x_1, \cdots, x_n). \tag{2.14}$$

Proof: First we write

$$\begin{aligned} &J(g)\sigma(x_{r_2} - x_{r_1}) \\ &= \frac{1}{2i}\sum_{k=1}^{n} g'(x_k)\sigma(x_{r_2} - x_{r_1}) \\ &\quad + \frac{1}{i}\left(g(x_{r_1})\frac{\partial}{\partial x_{r_1}} + g(x_{r_2})\frac{\partial}{\partial x_{r_2}}\right)\sigma(x_{r_2} - x_{r_1}). \end{aligned} \tag{2.15}$$

Setting $\xi = x_{r_1} + x_{r_2}$ and $\eta = x_{r_1} - x_{r_2}$, we find that the second term of (2.15) becomes

$$\begin{aligned} &\frac{1}{i}\left\{\left[g\left(\frac{\xi+\eta}{2}\right) + g\left(\frac{\xi-\eta}{2}\right)\right]\frac{\partial}{\partial\xi}\right. \\ &\qquad \left. + \left[g\left(\frac{\xi+\eta}{2}\right) - g\left(\frac{\xi-\eta}{2}\right)\right]\frac{\partial}{\partial\eta}\right\}\sigma(\eta), \end{aligned} \tag{2.16}$$

which vanishes since $\partial\sigma(\eta)/\partial\eta = 2\delta(\eta)$.

Thus (2.13) is proved.

Let us now write

$$\sigma(x_1, \cdots, x_n) = \sigma(x_{r_2} - x_{r_1})\hat{\sigma}(x_1, \cdots, x_n). \tag{2.17}$$

Using (2.13), one can show that (for $f \in \mathfrak{J}$)

$$\sigma(x_{r_2} - x_{r_1})J(g)f = J(g)[\sigma(x_{r_2} - x_{r_1})f] \tag{2.18}$$

by evaluating with respect to a general Schwartz-space function $k(x_1, \cdots, x_n)$. By Lemma 1, Eq. (2.18) makes sense in L^2.

In turn, suppose that $\hat{\sigma}(x_1, \cdots, x_n)f \in L^2$ has been shown to be in the domain of $J(g)^{**}$, where $\hat{\sigma}$ is the product of N terms of the form $\sigma(x_{r_2} - x_{r_1})$. Then the generalized function $J(g)[\sigma(x_{r_2} - x_{r_1})\hat{\sigma}f]$ equals

$$\sigma(x_{r_2} - x_{r_1})J(g)^{**}[\hat{\sigma}f],$$

since

$$\begin{aligned} &(J(g)[\sigma\hat{\sigma}f], k) \\ &\quad = (\hat{\sigma}f, \sigma J(g)k) = (\hat{\sigma}f, J(g)^{**}[\sigma k]) \\ &\quad = (\sigma J(g)^{**}[\hat{\sigma}f], k) \quad \text{for arbitrary} \quad k \in \mathfrak{J}. \end{aligned}$$

Letting $f(x_1, \cdots, x_n) = 1$ on the support of g, we prove (2.14) by induction on N. QED

Theorem 1: In the case of one spatial dimension, the n-particle representations of the current algebra (1.10)–(1.12) having different exchange symmetries are unitarily equivalent.

Proof: Define $Q: \mathcal{H}_s \to \mathcal{H}_a$ by $(Q\Psi)(x_1, \cdots, x_n) = \sigma(x_1, \cdots, x_n)\Psi$. Q is unitary, and $Q\rho_s(f) = \rho_a(f)Q$. Let $J(g)$ be the differential operator (2.12), and $J_s(g)^{**}$ or $J_a(g)^{**}$ its self-adjoint extension in $\mathcal{H}_s$ or $\mathcal{H}_a$. For $\Psi \in D_{J_s(g)^{**}}$ in H_s, the generalized function $J(g)[\sigma(x_1, \cdots, x_n)\Psi]$ equals $\sigma(x_1, \cdots, x_n)J_s(g)^{**}\Psi$, as may be shown by evaluating with respect to a general Schwartz-space function and using Lemma 2.

Then $\sigma(x_1, \cdots, x_n)\Psi \in \mathcal{H}_a$ is in the domain of $J_a(g)^{**}$.

Thus Q preserves the domain of $J(g)^{**}$, and we can conclude that $QJ_s(g) = J_a(g)Q$. QED

In Sec. 3, we shall see that Theorem 1 cannot be extended to the case of more than one spatial dimension; in that case, representations with different exchange symmetries are unitarily *in*equivalent.

3. EXPONENTIATION OF THE CURRENT ALGEBRA

We have seen that the smeared currents are apt to be unbounded operators and thus have domains of definition which are, in general, not all of $\mathcal{H}$. Our goal is to exponentiate the current algebras and obtain *groups*, which can be represented by unitary (hence bounded) operators. This procedure is analogous to replacing the relativistic equal-time canonical commutation relations of field theory by the Weyl group. We now turn our attention to performing the exponentiation for the nonrelativistic model reviewed in Sec. 1.

A most useful formula will be

$$e^A B e^{-A} = \sum_{n=0}^{\infty} (n!)^{-1}(\text{ad}^n A)B, \tag{3.1}$$

where A and B are operators and where

$$(\text{ad } A)B = [A, B]. \tag{3.2}$$

(3.1) may be proved formally by expanding e^A and e^{-A} in power series and regrouping the terms.

A. Vector Fields and Flows

Let $f \in \mathfrak{J}$ be a function of a single variable, and let

$$F(x) = \int_b^x \frac{dx'}{f(x')} \tag{3.3}$$

be defined in a region between zeros a_1 and a_2 of f, for an arbitrary fixed b in that region. a_1 and a_2 may be $\pm\infty$. Then $F: (a_1, a_2) \to \mathbb{R}$ and $F^{-1}: \mathbb{R} \to (a_1, a_2)$

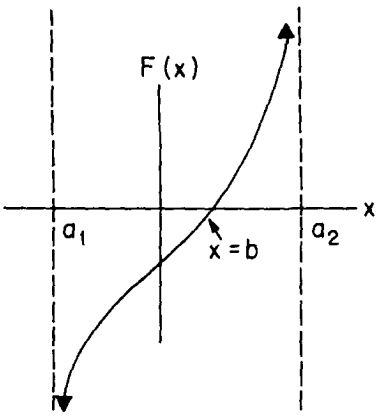

FIG. 1. The function $F(x)$ defined between zeros a_1 and a_2 of $f(x)$.

are one-to-one, onto, C_∞ functions (Fig. 1). F^{-1} is C_∞ since $F^{-1}(F(x)) = x$, $(F^{-1})'(F(x)) \cdot F'(x) = 1$, whence

$$(F^{-1})'(\xi) = 1/F'(F^{-1}(\xi)). \tag{3.4}$$

If F^{-1} is C_n, we can deduce from (3.4) that it is C_{n+1}; hence F^{-1} is C_∞.

Define

$$\varphi_t^f(x) = F^{-1}(F(x) + t)\big|_{\substack{x\in(a_1,a_2),\\ t\in\mathbb{R}}}, \tag{3.5}$$

and set

$$\varphi_t^f(a_1) = a_1, \quad \varphi_t^f(a_2) = a_2, \quad \forall\ t. \tag{3.6}$$

Then $\varphi_t^f(x)$ may be interpreted as the *flow for time t by the vector field f*. That is, $\varphi_t^f(x)$ satisfies the differential equation

$$\frac{\partial \varphi_t^f(x)}{\partial t} = f(\varphi_t^f(x)), \tag{3.7}$$

with the boundary condition

$$\varphi_{t=0}^f(x) = x. \tag{3.8}$$

In fact,

$$\begin{aligned}\frac{\partial \varphi_t^f(x)}{\partial t} &= (F^{-1})'(F(x) + t)\\ &= [F'(F^{-1}(F(x) + t))]^{-1}, \quad \text{by (3.4)},\\ &= [F'(\varphi_t^f(x))]^{-1}\\ &= f(\varphi_t^f(x)), \quad \text{by (3.3)}.\end{aligned}$$

In particular, $\varphi_t^f(x)$ is independent of our original choice of b.[14]

Next, consider the expression $\{\exp[tg(x)\,(d/dx)]\}f$, for $f, g \in \mathfrak{I}$. Setting $g(x)d/dx = d/d\xi$, we then have $d\xi/dx = 1/g$ and $\xi = G(x)$, where G is defined as in (3.3) in a region between zeros of g. Hence

$$\begin{aligned}\left[\exp\left(tg(x)\frac{d}{dx}\right)\right]f(x) &= \exp\left(t\frac{d}{d\xi}\right)f(G^{-1}(\xi))\\ &= f(G^{-1}(\xi + t))\\ &= f(G^{-1}(G(x) + t)).\end{aligned}$$

But we saw in (3.5)–(3.8) that $\varphi_t^g(x) = G^{-1}(G(x) + t)$ defines the flow for time t by the vector field g.

We are thus led to the important equation

$$e^{t\mathbf{g}\cdot\nabla}f(\mathbf{x}) = f(\boldsymbol{\varphi}_t^{\mathbf{g}}(\mathbf{x})), \tag{3.9}$$

where $\boldsymbol{\varphi}_t^{\mathbf{g}}(\mathbf{x})$ satisfies

$$\frac{\partial \boldsymbol{\varphi}_t^{\mathbf{g}}(\mathbf{x})}{\partial t} = \mathbf{g}(\boldsymbol{\varphi}_t^{\mathbf{g}}(\mathbf{x})) \tag{3.10}$$

and $\boldsymbol{\varphi}_{t=0}^{\mathbf{g}}(\mathbf{x}) = \mathbf{x}$. Equation (3.9) may be verified directly by differentiating with respect to t, at $t = 0$.[15]

B. Obtaining the Group

The nonrelativistic model is defined by Eqs. (1.10)–(1.12). Of course, (1.10) immediately gives us

$$e^{i\rho(f)}e^{i\rho(g)} = e^{i\rho(f+g)}. \tag{3.11}$$

From (1.11), using (3.1), we have

$$\begin{aligned}e^{iJ(\mathbf{g})}\rho(f)e^{-iJ(\mathbf{g})} &= \sum_{n=0}^{\infty}\frac{i^n}{n!}[\mathrm{ad}^n J(\mathbf{g})]\rho(f)\\ &= \sum_{n=0}^{\infty}\rho\left(\frac{(\mathbf{g}\cdot\nabla)^n}{n!}f\right) = \rho(e^{\mathbf{g}\cdot\nabla}f),\end{aligned} \tag{3.12}$$

where, in the last step, ρ is taken to be in some sense continuous in its argument. Thus,

$$e^{iJ(\mathbf{g})}e^{i\rho(f)} = \exp[i\rho(e^{\mathbf{g}\cdot\nabla}f)]e^{iJ(\mathbf{g})}. \tag{3.13}$$

Also,

$$\begin{aligned}e^{i\rho(f)}J(\mathbf{g})e^{-i\rho(f)} &= \sum_{n=0}^{\infty}\frac{i^n}{n!}[\mathrm{ad}^n \rho(f)]J(\mathbf{g})\\ &= J(\mathbf{g}) - \rho(\mathbf{g}\cdot\nabla f);\end{aligned} \tag{3.14}$$

thus

$$e^{i\rho(f)}e^{iJ(\mathbf{g})} = e^{i[J(\mathbf{g})-\rho(\mathbf{g}\cdot\nabla f)]}e^{i\rho(f)}. \tag{3.15}$$

Equation (3.1) does not give us the means to multiply $\exp[itJ(\mathbf{f})]$ by $\exp[itJ(\mathbf{g})]$. Let us look at this question in the 1-particle, 1-dimensional Fock representation. Here,

$$J(f) = -if\frac{d}{dx} - \tfrac{1}{2}if'$$

and

$$\exp[iJ(f)] = \exp\left(f\frac{d}{dx} + \tfrac{1}{2}\rho(f')\right),$$

or

$$e^{i[J(f)-\rho(-\frac{1}{2}if')]} = e^{f(d/dx)}. \tag{3.16}$$

But, by (3.15),

$$e^{i[J(f)-\rho(-\frac{1}{2}if')]} = e^{i\rho(h)}e^{iJ(f)}e^{-i\rho(h)}, \tag{3.17}$$

where $f(x)(dh/dx) = -\frac{1}{2}if'$ or $h = \frac{1}{2}i\ln(f)$. Thus

$$e^{iJ(f)} = e^{-\frac{1}{2}\ln f}e^{f(d/dx)}e^{+\frac{1}{2}\ln f} \tag{3.18}$$

and, from (3.9),

$$e^{itJ(f)}\Psi(x) = \Psi(\varphi_t^f(x))\left(\frac{f(\varphi_t^f(x))}{f(x)}\right)^{\frac{1}{2}}. \quad (3.19)$$

Since one can easily show [e.g., from (3.5)] that

$$\frac{f(\varphi_t^f(x))}{f(x)} = \frac{d}{dx}\,\varphi_t^f(x), \quad (3.20)$$

we obtain

$$e^{itJ(f)}\Psi(x) = \Psi(\varphi_t^f(x))\left(\frac{d}{dx}\,\varphi_t^f(x)\right)^{\frac{1}{2}} \quad (3.21)$$

in the 1-particle, 1-dimensional Fock representation. (3.21) clearly defines a unitary operator, as is desired.

Taking note that $\exp[itJ(f)]$ does not depend directly on f, but on φ_t^f, let us define, in general,

$$U(f) = e^{i\rho(f)} \quad (3.22)$$

and

$$V(\boldsymbol{\varphi}_t^{\mathbf{g}}) = e^{it,J(\mathbf{g})}. \quad (3.23)$$

Equation (3.21), in the higher-dimensional Fock case, becomes

$$V(\boldsymbol{\varphi}_t^{\mathbf{g}})\Psi(\mathbf{x}) = \Psi(\boldsymbol{\varphi}_t^{\mathbf{g}}(\mathbf{x}))\left[\det\left(\frac{\partial(\varphi_t^{\mathbf{g}})^k}{\partial x^j}\right)\right]^{\frac{1}{2}}, \quad (3.24)$$

where we recognize the square root of the Jacobian as just that factor necessary to make V unitary.

Finally, it is not hard to show in the Fock representation that

$$V(\boldsymbol{\varphi})V(\boldsymbol{\psi}) = V(\boldsymbol{\psi} \circ \boldsymbol{\varphi}) \quad (3.25)$$

for flows $\boldsymbol{\psi}$ and $\boldsymbol{\varphi}$.

Let us summarize these results as a theorem.

Theorem 2: Let $f \in \mathfrak{J}$, and let $\mathbf{g}$ be a C_∞ vector field on $\mathbb{R}^s$ with components in $\mathfrak{J}$. The flow $\boldsymbol{\varphi}_t^{\mathbf{g}}$ is then C_∞ in $\mathbf{x}$ and t and may be said to suitably approximate the identity mapping as $|\mathbf{x}| \to \infty$. Let $\boldsymbol{\psi}$ be a C_∞ diffeomorphism from $\mathbb{R}^s$ onto $\mathbb{R}^s$, which is a finite product of flows $\boldsymbol{\varphi}_t^{\mathbf{g}}$. For $\Psi \in L^2(\mathbb{R}^s)$, define

$$U(f)\Psi(\mathbf{x}) = e^{if(\mathbf{x})}\Psi(\mathbf{x})$$

and

$$V(\boldsymbol{\psi})\Psi(\mathbf{x}) = \Psi(\boldsymbol{\psi}(\mathbf{x}))\left[\det\left(\frac{\partial\psi^k}{\partial x^j}\right)\right]^{\frac{1}{2}}.$$

Then $U(tf)$ and $V(\boldsymbol{\varphi}_t^{\mathbf{g}})$ are strongly continuous 1-parameter unitary groups. The infinitesimal generators of these groups, which exist and define self-adjoint operators by Stone's theorem,[16] are, respectively, $\rho(f)$ and $J(\mathbf{g})$, where $\rho(f)\Psi(\mathbf{x}) = f(\mathbf{x})\Psi(\mathbf{x})$ and $J(\mathbf{g})\Psi(\mathbf{x}) = -\frac{1}{2}i(\mathbf{g}\cdot\boldsymbol{\nabla} + \boldsymbol{\nabla}\cdot\mathbf{g})\Psi(\mathbf{x})$. That is, $U(tf) = e^{it\rho(f)}$ and $V(\boldsymbol{\varphi}_t^{\mathbf{g}}) = e^{itJ(\mathbf{g})}$. U and V satisfy the multiplication rules

$$U(f)U(g) = U(f+g), \quad (3.26)$$

$$V(\boldsymbol{\psi})U(f) = U(f\circ\boldsymbol{\psi})V(\boldsymbol{\psi}), \quad (3.27)$$

$$V(\boldsymbol{\varphi})V(\boldsymbol{\psi}) = V(\boldsymbol{\psi}\circ\boldsymbol{\varphi}). \quad (3.28)$$

Proof: All of the assertions are directly verifiable by computation. In particular,

$$\rho(f)\Psi = \lim_{t\to 0}\frac{U(tf)-I}{it}\Psi \quad (3.29)$$

and

$$J(\mathbf{g})\Psi = \lim_{t\to 0}\frac{V(\boldsymbol{\varphi}_t^{\mathbf{g}})-I}{it}\Psi, \quad (3.30)$$

where the limits are understood in the L^2 sense. We omit the details. QED

C. Remarks

(1) Henceforth, a representation of the group defined by Eqs. (3.26)–(3.28) will be taken as the basis of the theory. We should note that not every representation of a Lie algebra leads to a corresponding representation of the Lie group. In fact, Nelson gives us an example of two symmetric operators in a Hilbert space $\mathcal{H}$, defined and commuting on a common, dense, invariant domain D of essential self-adjointness, whose spectral resolutions do not commute.[17] In order to ensure that a particular representation of the unexponentiated formalism leads to one of the exponentiated version, one needs to postulate analytic vectors in the domain D. We will not go into this subject further here. In Sec. 5, sufficient conditions are developed for recovering the infinitesimal generators in a representation of the exponentiated formalism.

(2) We saw in Theorem 2 that, in the Fock representation, the currents can be exponentiated with $U(tf) = e^{it\rho(f)}$ and $V(\boldsymbol{\varphi}_t^{\mathbf{g}}) = e^{itJ(\mathbf{g})}$, so that U and V satisfy

$$U(f_1)V(\boldsymbol{\psi}_1)U(f_2)V(\boldsymbol{\psi}_2) = U(f_1 + f_2\circ\boldsymbol{\psi}_1)V(\boldsymbol{\psi}_2\circ\boldsymbol{\psi}_1). \quad (3.31)$$

But (3.31) is the multiplication law for elements of the *semidirect product* $\mathfrak{J} \wedge \mathcal{K}$ of the groups defined by the set $\mathfrak{J}$ of all f's (under addition) and the set $\mathcal{K}$ of all $\boldsymbol{\psi}$'s (under composition). The mapping $\mathfrak{J} \times \mathcal{K} \to \mathfrak{J}$, which defines the semidirect product, is given by $(f, \boldsymbol{\psi}) \to f\circ\boldsymbol{\psi}$.

The additive group of the f's has already been identified as Schwartz space $\mathfrak{J}$.[18] Among its advantages, $\mathfrak{J}$ is suitable for the representation theory discussed in Sec. 4 and is a *normal* subgroup of $\mathfrak{J} \wedge \mathcal{K}$.

In Ref. 5, a definition and a topology are proposed for $\mathcal{K}$, such that the group operations of $\mathcal{J} \wedge \mathcal{K}$ are jointly continuous. ($\mathcal{J}$ is taken to have the usual Schwartz-space topology.) The 1-parameter subgroups φ_t^g in $\mathcal{K}$ are continuous in t. Thus $\mathcal{J} \wedge \mathcal{K}$ defines a topological group, and, if we choose to look at its strongly continuous representations, we are assured that $U(tf)$ and $V(\varphi_t^g)$ define strongly continuous 1-parameter groups of unitary operators.

For simplicity, we shall often restrict our attention to elements of $\mathcal{K}$ with *compact support*, i.e., satisfying $\psi(\mathbf{x}) = \mathbf{x}$ outside of a compact set.

D. The Nonrelativistic Model with Spin

This model is defined by Eqs. (1.10)–(1.12) together with (1.22)–(1.24). Of course, the ρ-J commutators give us the group $\mathcal{J} \wedge \mathcal{K}$, as before. Since ρ and Σ commute, we also have immediately

$$e^{i\Sigma(\mathbf{g})}e^{i\rho(f)} = e^{i\rho(f)}e^{i\Sigma(\mathbf{g})}. \tag{3.32}$$

Applying (3.1) to the Σ-J commutator, we obtain

$$\begin{aligned} e^{iJ(\mathbf{g})}\Sigma(\mathbf{h})e^{-iJ(\mathbf{g})} &= \sum_{n=0}^{\infty} \frac{i^n}{n!}[\mathrm{ad}^n J(\mathbf{g})]\Sigma(\mathbf{h}) \\ &= \sum_{n=0}^{\infty} \frac{1}{n!}\Sigma((\mathbf{g}\cdot\nabla)^n\mathbf{h}) = \Sigma(e^{\mathbf{g}\cdot\nabla}h) \end{aligned} \tag{3.33}$$

or, recalling (3.9),

$$e^{itJ(\mathbf{g})}e^{i\Sigma(\mathbf{h})}e^{-itJ(\mathbf{g})} = \exp\,[i\Sigma(\mathbf{h}\circ\varphi_t^{\mathbf{g}})]. \tag{3.34}$$

Applying (3.1) to the Σ-Σ commutator,

$$\begin{aligned} e^{i\Sigma(\mathbf{f})}\Sigma(\mathbf{g})e^{-i\Sigma(\mathbf{f})} &= \Sigma(\mathbf{g}) + i^2\Sigma(\mathbf{f}\times\mathbf{g}) + (i^4/2!)\Sigma(\mathbf{f}\times(\mathbf{f}\times\mathbf{g})) \\ &\quad + (i^6/3!)\Sigma(-|\mathbf{f}|^2(\mathbf{f}\times\mathbf{g})) \\ &\quad + (i^8/4!)\Sigma(-|\mathbf{f}|^2(\mathbf{f}\times(\mathbf{f}\times\mathbf{g}))) + \cdots, \end{aligned} \tag{3.35}$$

where we have used the identity $\mathbf{f}\times(\mathbf{f}\times\mathbf{g}) = (\mathbf{f}\cdot\mathbf{g})f - (\mathbf{f}\cdot\mathbf{f})\mathbf{g}$, whence $\mathbf{f}\times[\mathbf{f}\times(\mathbf{f}\times\mathbf{g})] = -|\mathbf{f}|^2\cdot(\mathbf{f}\times\mathbf{g})$. (3.35) becomes

$$\Sigma(\mathbf{g} - (\mathbf{f}\times\mathbf{g})\sin|\mathbf{f}| - \hat{f}\times(\hat{f}\times\mathbf{g})\cdot(\cos|\mathbf{f}| - 1)),$$

where $\hat{f} = \mathbf{f}/|\mathbf{f}|$. But, if $R(\hat{n}, \theta)\mathbf{y}$ denotes the vector obtained by rotating the vector $\mathbf{y}$ about the unit vector $\hat{n}$ by angle θ, one can show that

$$\mathbf{g} - (\hat{f}\times\mathbf{g})\sin|\mathbf{f}| - \hat{f}\times(\hat{f}\times\mathbf{g})(\cos|\mathbf{f}| - 1) = R(\hat{f}, -|\mathbf{f}|)\mathbf{g}, \tag{3.36}$$

using

$$R(\hat{n}, \theta)\mathbf{y} = (\cos\theta)\mathbf{y} + (1 - \cos\theta)\hat{n}\hat{n}\cdot\mathbf{y} + (\sin\theta)\hat{n}\times\mathbf{y} \tag{3.37}$$

and

$$\hat{f}\times(\hat{f}\times\mathbf{g}) = \hat{f}(\mathbf{g}\cdot\hat{f}) - \mathbf{g}.$$

Thus we obtain

$$e^{i\Sigma(\mathbf{f})}e^{i\Sigma(\mathbf{g})}e^{-i\Sigma(\mathbf{f})} = \exp\,[-i\Sigma(R(\hat{f}, -|\mathbf{f}|)\mathbf{g})]. \tag{3.38}$$

This equation suggests that the generators of the group of exponentiated Σ's should be *local rotations* $R(\mathbf{x})$ with, say, compact support, that is, C_∞ mappings $R: \mathbb{R}^3 \to SO_3$ with $R(\mathbf{x}) = I$ outside a compact set. The law for group multiplication should be $(R_1R_2)(\mathbf{x}) = R_1(\mathbf{x})R_2(\mathbf{x})$, where the right-hand side denotes multiplication in SO_3. We would then have the correspondence $e^{i\Sigma(\mathbf{f})} = R_{\mathbf{f}}(\mathbf{x})$ where

$$R_{\mathbf{f}}(\mathbf{x}) = \widehat{R(f(\mathbf{x}), -|\mathbf{f}(\mathbf{x})|)}.$$

But, as is well known, if we limit ourselves to the rotation group, we will be able to obtain only representations in which the particles have integral spin. In order to accommodate half-integral spins, we must replace the rotation group by SU_2. Therefore, we let $T: \mathbb{R}^3 \to SU_2$ be a C_∞ mapping with compact support and $\mathcal{G}$ be the group of all T's under the multiplication law $(T_1T_2)(\mathbf{x}) = T_1(\mathbf{x})T_2(\mathbf{x})$. We then have the correspondence $e^{i\Sigma(\mathbf{f})} = T_{\mathbf{f}}$, where

$$T_{\mathbf{f}}(\mathbf{x}) = \exp\,[-\tfrac{1}{2}i\mathbf{f}(\mathbf{x})\cdot\boldsymbol{\sigma}].$$

Here, $\boldsymbol{\sigma} = (\sigma_1, \sigma_2, \sigma_3)$ are the Pauli spin matrices satisfying (1.16).

We observe that

$$e^{-\frac{1}{2}i\mathbf{f}\cdot\boldsymbol{\sigma}}e^{-\frac{1}{2}i\mathbf{g}\cdot\boldsymbol{\sigma}}e^{+i\mathbf{f}\cdot\frac{1}{2}\boldsymbol{\sigma}} = \exp\,[-\tfrac{1}{2}iR(\hat{f}, -|\mathbf{f}|)\mathbf{g}\cdot\boldsymbol{\sigma}] \tag{3.39}$$

since

$$[\tfrac{1}{2}\mathbf{f}\cdot\boldsymbol{\sigma}, \tfrac{1}{2}\mathbf{g}\cdot\boldsymbol{\sigma}] = \tfrac{1}{2}i(\mathbf{f}\times\mathbf{g})\cdot\boldsymbol{\sigma}. \tag{3.40}$$

Let $\mathfrak{D}^{(j)}(M)$ be the $(2j+1)$-dimensional representation of SU_2, where $M \in SU_2$. Let $\Psi(\mathbf{x})$ be a $(2j+1)$-component spinor. Define the representation W of $\mathcal{G}$ by

$$W(T)\Psi(\mathbf{x}) = \mathfrak{D}^{(j)}(T(\mathbf{x}))\Psi(\mathbf{x}). \tag{3.41}$$

Equation (3.41), along with $U(f)\Psi(\mathbf{x}) = e^{if(\mathbf{x})}\Psi(\mathbf{x})$ and

$$V(\psi)\Psi(\mathbf{x}) = \Psi(\psi(\mathbf{x}))\det\left[\left(\frac{\partial\psi^k}{\partial x^j}(\mathbf{x})\right)\right]^{\frac{1}{2}},$$

defines the 1-particle Fock representation having spin j. Equations (3.26)–(3.28), together with

$$W(T)U(f) = U(f)W(T), \tag{3.42}$$

$$V(\psi)W(T) = W(T\circ\psi)V(\psi), \tag{3.43}$$

$$W(T_1)W(T_2) = W(T_1T_2), \tag{3.44}$$

are satisfied.

In short, $U(f)W(T)V(\psi)$ defines a representation of $(\mathfrak{I} \otimes \mathfrak{G}) \wedge \mathcal{K}$, where $\otimes$ is the direct product and where $\wedge$ is the semidirect product defined by the law $\psi\colon (f, T) \to (f \circ \psi, T \circ \psi)$.

Let us remark that nonrelativistic quantum mechanics is supposed to be *Euclidean invariant*; thus there is also a unitary representation $E(\mathbf{a}, R)$ of the Euclidean group $\mathbb{R}^3 \wedge SO_3$ in the Hilbert space $\mathcal{K}$. E must satisfy

$$E(\mathbf{a}, R)U(f)E(\mathbf{a}, R)^{-1} = U(f_{(\mathbf{a},R)}), \tag{3.45}$$

where

$$f_{(\mathbf{a},R)}(\mathbf{x}) = f(R^{-1}(\mathbf{x} - \mathbf{a})). \tag{3.46}$$

The action of E on $\mathcal{K}$ is given by

$$E(\mathbf{a}, R)V(\psi)E(\mathbf{a}, R)^{-1} = V(\psi^{(\mathbf{a},R)}), \tag{3.47}$$

where

$$\psi^{(\mathbf{a},R)}(\mathbf{x}) = R\psi(R^{-1}(\mathbf{x} - \mathbf{a})) + \mathbf{a}. \tag{3.48}$$

It is easy to show that if $\psi = \varphi_t^{\mathbf{f}}$, then

$$\psi^{(\mathbf{a},R)} = \varphi_t^{(\mathbf{f}_{(\mathbf{a},R)})}. \tag{3.49}$$

Finally, when spin is included, we have

$$E(\mathbf{a}, R)W(T)E(\mathbf{a}, R)^{-1} = W(T_{(\mathbf{a},R)}), \tag{3.50}$$

where

$$T_{(\mathbf{a},R)}(\mathbf{x}) = T(R^{-1}(\mathbf{x} - \mathbf{a})). \tag{3.51}$$

E. The n-Particle Representation

The n-particle representation $U(f)V(\psi)$ of the group $\mathfrak{I} \wedge \mathcal{K}$ is given by

$$U(f)\Psi(\mathbf{x}_1, \cdots, \mathbf{x}_n) = \exp\left(i\sum_{k=1}^{n} f(\mathbf{x}_k)\right)\Psi(\mathbf{x}_1, \cdots, \mathbf{x}_n), \tag{3.52}$$

$$V(\psi)\Psi(\mathbf{x}_1, \cdots, \mathbf{x}_n) = \Psi(\psi(\mathbf{x}_1), \cdots, \psi(\mathbf{x}_n)) \prod_{k=1}^{n} \left[\det\left(\frac{\partial \psi^l}{\partial x^j}(\mathbf{x}_k)\right)\right]^{\frac{1}{2}}. \tag{3.53}$$

Just as ρ and J preserve the exchange symmetry of Ψ in the unexponentiated version (2.10) and (2.11), likewise U and V now preserve exchange symmetry.

Lemma 3: Let $\mathbf{g}$ be a C_∞ vector field with compact support. Then in the n-particle representation of $\mathfrak{I} \wedge \mathcal{K}$, $J(\mathbf{g})$ defines an essentially self-adjoint operator on the domain D of Schwartz-space functions in L^2, $\mathcal{H}_s$, or $\mathcal{H}_a$, respectively.

Proof[19]*:* With $V(\varphi_t^{\mathbf{g}}) = e^{itJ(\mathbf{g})}$, the domain D is invariant under $V(\varphi_t^{\mathbf{g}})$ in L^2, $\mathcal{H}_s$, or $\mathcal{H}_a$, respectively, as is obvious from (3.53). Suppose that $J(\mathbf{g})$ is not essentially self-adjoint on D. Then there exists a vector Ψ in the Hilbert space, $\Psi \neq 0$, such that, for all $\Phi \in D$, $(\Psi, J(\mathbf{g})\Phi) = \pm i(\Psi, \Phi)$. Now $(\Psi, V(\varphi_t^{\mathbf{g}})\Phi)$ is differentiable in t at $t = 0$; therefore, it is differentiable for all t by the invariance of D under $V(\varphi_t^{\mathbf{g}})$. Hence

$$-i\frac{\partial}{\partial t}(\Psi, V(\varphi_t^{\mathbf{g}})\Phi) = (\Psi, J(\mathbf{g})V(\varphi_t^{\mathbf{g}})\Phi) = \pm i(\Psi, V(\varphi_t^{\mathbf{g}})\Phi). \tag{3.54}$$

This differential equation has the solution

$$(\Psi, V(\varphi_t^{\mathbf{g}})\Phi) = (\Psi, \Phi)e^{\mp t}. \tag{3.55}$$

But $(\Psi, V(\varphi_t^{\mathbf{g}})\Phi)$ is a bounded function of t, while $(\Psi, \Phi)e^{\mp t}$ is unbounded, unless $(\Psi, \Phi) = 0$. Thus $(\Psi, \Phi) = 0$ for all Φ in D, and $\Psi = 0$ since D is dense, a contradiction. QED

This is the result promised in Lemma 2 and used in proving Theorem 1.

Theorem 3[20]*:* In the case of more than one spatial dimension, the n-particle representations of the current algebra (1.10)–(1.12) having different exchange symmetries are unitarily *in*equivalent.

Proof: Suppose that $Q\colon \mathcal{H}_s \to \mathcal{H}_a$ defines a unitary equivalence. From the fact that $QU_s(f) = U_a(f)Q$, where the subscripts refer to $\mathcal{H}_s$ or $\mathcal{H}_a$, we conclude that Q merely multiplies by a measurable, antisymmetric function $\sigma(\mathbf{x}_1, \cdots, \mathbf{x}_n)$. Since Q is unitary, $|\sigma(\mathbf{x}_1, \cdots, \mathbf{x}_n)| = 1$ almost everywhere.

From the fact that $QV_s(\psi) = V_a(\psi)Q$, we have $(\forall\, \psi \in \mathcal{K})$ that $\sigma(\psi(\mathbf{x}_1), \cdots, \psi(\mathbf{x}_n)) = \sigma(\mathbf{x}_1, \cdots, \mathbf{x}_n)$ almost everywhere. The idea of the proof is to find a flow ψ which, roughly speaking, exchanges $\mathbf{x}_1$ and $\mathbf{x}_2$, in the case of more than one spatial dimension. Then the invariance of σ under composition with ψ will contradict the antisymmetry of σ under coordinate exchange.

The details are as follows, for n particles in s spatial dimensions. Consider the set X^+ of all points $(\mathbf{x}_1, \cdots, \mathbf{x}_n)$ such that the complex number $\sigma(\mathbf{x}_1, \cdots, \mathbf{x}_n)$ is in the interior of some fixed half-plane H^+. Since σ is measurable, X^+ is a measurable set, in Lebesgue measure μ. Let Y be the set of all points $(\mathbf{x}_1, \cdots, \mathbf{x}_n)$ such that $|\mathbf{x}_1| < 1$, $|\mathbf{x}_2| < 1$, $|\mathbf{x}_3|$, $|\mathbf{x}_4|, \cdots, |\mathbf{x}_n| > 2$. In the definition of X^+, the half-plane H^+ may be chosen so that $\mu(X^+ \cap Y) > 0$.

For any measurable set X in $\mathbb{R}^m$ with $\mu(X) > 0$ and for any $\delta > 0$, one can find an open set $\mathcal{O}$ with $\mu(\mathcal{O}) < \infty$ such that $\mu(\mathcal{O} \cap X) > (1 - \delta)\mu(\mathcal{O})$, i.e., *most* of $\mathcal{O}$ is contained in X.[21]

Let $\mathcal{O}$, then, be an open set in $\mathbb{R}^{sn} \cap Y$, such that most of $\mathcal{O}$ is in $X^+ \cap Y$.

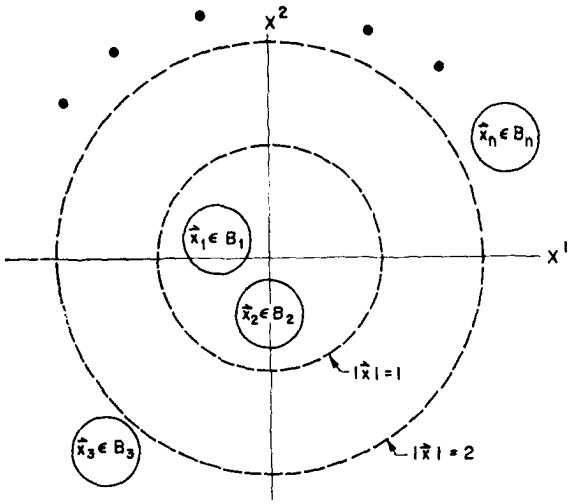

FIG. 2. Proof of Theorem 3.

Let $B_1, \cdots, B_n$ denote open balls in $\mathbb{R}^s$. Product sets of the form $B_1 \times \cdots \times B_n$, where all B_j have equal radii, form a base for the topology in $\mathbb{R}^{sn}$. It is not difficult to show that any open set in $\mathbb{R}^{sn}$ can be written as the union of countably many *disjoint* sets of the form $B_1 \times \cdots \times B_n$, where all B_j have equal radii, together with a set of measure zero.

Therefore, one can find $B = B_1 \times \cdots \times B_n \subset Y$, where all B_j have equal radii, such that most of B is in $X^+ \cap Y$. (See Fig. 2.)

Let ψ be a flow on $\mathbb{R}^s$ with support in the closed ball of radius 2 about the origin. ψ may be chosen to be volume-preserving inside the closed ball of radius 1 about the origin, so that $\psi(B_1) = B_2$, $\psi(B_2) = B_1$. Inside the closed ball of radius 1, ψ is simply rotation about an axis or axes drawn perpendicular to the line joining the centers of B_1 and B_2 at the midway point between them. Of course, ψ only exists if $s > 1$.

Now σ takes on values inside H^+ for $(\mathbf{x}_1, \cdots, \mathbf{x}_n) \in X^+$; hence σ maps *most* of B into H^+. As σ is antisymmetric with respect to exchange of $\mathbf{x}_1$ and $\mathbf{x}_2$, σ maps most of $B' = B_2 \times B_1 \times \cdots \times B_n$ into H^-, the complementary half-plane. But $\psi: B \to B'$ is volume-preserving, and, if σ is invariant under ψ, σ must map most of B' into H^+, which is impossible.

QED

4. REPRESENTATION THEORY

For several reasons, it is important to study the different representations of the groups obtained in Sec. 3, with the aim of classifying them.

(1) We have seen how the current algebra can be obtained from the nonrelativistic canonical commutation or anticommutation relations of field theory, both formally and in the Fock representation. The question remains open whether *every* representation of the current algebra arises, like the Fock, from an underlying representation of the canonical formalism. In a sense, this question is fundamental to the whole approach, for it asks whether we can do anything with currents which we cannot do with fields. It becomes even more significant in relation to relativistic current algebras.

(2) We have succeeded in identifying representations of $\mathfrak{J} \wedge \mathfrak{K}$ which describe n identical particles and in demonstrating the unitary inequivalence of n-particle representations whose states obey different statistics, in the case of more than one spatial dimension. In a later paper, representations will be obtained which describe infinitely many identical particles, at a constant average density.[22] It is important to have a unified framework in which all of these representations can be studied.

(3) Grodnik and Sharp examine a volume-cutoff theory of n particles.[23] They obtain a formal representation of the current algebra through multiplication and differentiation on a space of functions of infinitely many complex variables. The theory studied in this section can be applied to give rigorous meaning to such a representation.

(4) It is certainly of physical interest to inquire whether the current algebra can be recovered (on an appropriate domain) by taking the infinitesimal generators of the continuous 1-parameter subgroups, in a given representation of the exponentiated formalism. In Sec. 5, we shall characterize some representations in which this is indeed the case.

A. Formalisms

A representation of one of the groups from Sec. 3 certainly involves a representation $U(f)$ of the Abelian subgroup $\mathfrak{J}$. The idea common to all of the methods we discuss is to express these $U(f)$ as diagonal operators. We restrict ourselves initially to representations in which there exists a cyclic vector Ω for the C^*-algebra associated with the $U(f)$.

Let $\mathfrak{A}$ be the C^*-algebra generated by $\{U(f) \mid f \in \mathfrak{J}\}$.[24] The *Gel'fand spectral theory* of commutative C^*-algebras allows us to realize cyclic representations of $\mathfrak{A}$ by multiplication operators on $L^2_\mu(Z)$, where Z is the spectrum of $\mathfrak{A}$ and where μ is a basic measure on Z.[25,5]

In fact, let $U(f)$ be a strongly continuous unitary representation of $\mathfrak{J}$ in the Hilbert space $\mathfrak{H}$, and let $\Omega \in \mathfrak{H}$ be a normalized cyclic vector for the representation. Let

$$Z = \{\zeta: U(\mathfrak{J}) \to \mathbb{C};\ (\forall f, g \in \mathfrak{J}),\ |\zeta(U(f))| = 1,\ \text{and}\ \zeta(U(f))\zeta(U(g)) = \zeta(U(f+g))\}.$$

The points of Z are in one-to-one correspondence with the spectrum of $\mathfrak{A}$ and with the set of (not necessarily continuous) linear mappings $\ell_\zeta: \mathfrak{I} \to \mathbb{R} (\mathrm{mod}\ 2\pi)$ according to the formula

$$\zeta(U(f)) = e^{i(\ell_\zeta, f)}. \tag{4.1}$$

Then $\mathcal{H}$ can be realized as $L^2_\mu(Z)$, where the basic measure μ satisfies

$$(\Omega, U(f)\Omega) = \int_Z e^{i(\ell_\zeta, f)}\, d\mu(\zeta). \tag{4.2}$$

Ω can be represented by $\Omega(\zeta) \equiv 1$, and, for all $\Phi \in L^2_\mu(Z)$,

$$[U(f)\Phi](\zeta) = e^{i(\ell_\zeta, f)}\Phi(\zeta). \tag{4.3}$$

Suppose next that we have a unitary representation $U(f)V(\Psi)$ of $\mathfrak{I} \wedge \mathfrak{K}$ in $\mathcal{H}$, with $\Omega \in \mathcal{H}$ cyclic for $\mathfrak{A}$. For $\psi \in \mathfrak{K}$, we can define an action of ψ on Z by setting $(\psi^*\zeta)[U(f)] = \zeta(U(f \circ \psi))$. Let $\psi^*\ell_\zeta = \ell_{\psi^*\zeta}$, so that $(\psi^*\ell_\zeta, f) = (\ell_\zeta, f \circ \psi)$. Clearly

$$(\psi_2 \circ \psi_1)^*\zeta = \psi_2^*\psi_1^*\zeta.$$

Now, for $\psi \in \mathfrak{K}$, $V(\psi)$ is a unitary operator in $\mathcal{H}$. Let $\Omega_1 = V(\psi)\Omega$; then Ω_1 is also a cyclic vector for $\mathfrak{A}$, and the measure μ_1, defined with respect to Ω_1 to satisfy (4.2), is a second basic measure on Z.

It follows from the existence of the unitary $V(\psi)$ that μ_1 and μ are equivalent measures. Furthermore,

$$\begin{aligned}\int_Z \exp[i(\ell_\zeta, f)]\, d\mu_1(\zeta) &= (\Omega_1, U(f)\Omega_1)\\ &= (\Omega, U(f \circ \psi^{-1})\Omega)\\ &= \int_Z \exp[i(\ell_{\psi^{-1}*\zeta}, f)]\, d\mu(\zeta)\\ &= \int_Z \exp[i(\ell_\zeta, f)]\, d\mu(\psi^*\zeta),\end{aligned} \tag{4.4}$$

whence

$$d\mu_1(\zeta) = d\mu(\psi_1^*\zeta). \tag{4.5}$$

Since μ and μ_1 are equivalent, we can write

$$d\mu_1(\zeta) = \frac{d\mu_1}{d\mu}(\zeta)\, d\mu(\zeta), \tag{4.6}$$

where $(d\mu_1/d\mu)(\zeta)$ denotes the Radon–Nikodym derivative.[26] A measure μ satisfying the property that, for any measurable set X, $\mu(X) = 0 \Leftrightarrow \mu(\psi^*X) = 0$, $\forall \psi \in \mathfrak{K}$, will be called *quasi-invariant for* $\mathfrak{K}$.

Using the fact that $U(f)\Omega$ spans a dense subspace of $\mathcal{H}$, we easily show that, for $\Psi \in L^2_\mu(Z)$,

$$[V(\psi)\Psi](\zeta) = \chi_\psi(\zeta)\Psi(\psi^*\zeta)\left(\frac{d\mu(\psi^*\zeta)}{d\mu(\zeta)}\right)^{\frac{1}{2}}, \tag{4.7}$$

where the "*multiplier*" $\chi_\psi(\zeta)$ is a measurable complex-valued function of modulus one, which depends on ψ.[27] One can verify directly that (4.3) and (4.7) define a representation of $\mathfrak{I} \wedge \mathfrak{K}$, where, in order that $V(\psi_2)V(\psi_1) = V(\psi_1 \circ \psi_2)$, the $\chi_\psi(\zeta)$ must satisfy, $\forall\, \psi_1, \psi_2$,

$$\chi_{\psi_2}(\zeta)\chi_{\psi_1}(\psi_2^*\zeta) = \chi_{\psi_1 \circ \psi_2}(\zeta) \tag{4.8}$$

almost everywhere. The choice $\chi_\psi(\zeta) \equiv 1$, $\forall\, \psi$, will always satisfy (4.8), though it may define a representation inequivalent to the one with which we started.

In the noncyclic case, $L^2_\mu(Z)$ is replaced by $\int_Z \mathcal{H}_\zeta\, d\mu(\zeta)$, with $\dim \mathcal{H}_\zeta \neq 1$. $\Psi \in \mathcal{H}$ corresponds to the field of vectors $\langle \Psi(\zeta) \rangle$, with $\Psi(\zeta) \in \mathcal{H}_\zeta$. Then

$$\langle (U(f)\Psi)(\zeta) \rangle = \langle \exp[i(\ell_\zeta, f)]\Psi(\zeta) \rangle \tag{4.9}$$

and

$$\langle (V(\psi)\Psi)(\zeta) \rangle = \left\langle \chi_{\psi^*\zeta}(\zeta)\Psi(\psi^*\zeta)\left(\frac{d\mu(\psi^*\zeta)}{d\mu(\zeta)}\right)^{\frac{1}{2}} \right\rangle, \tag{4.10}$$

where the "multiplier"

$$\chi_{\psi^*\zeta}(\zeta): \mathcal{H}_{\psi^*\zeta} \to \mathcal{H}_\zeta \tag{4.11}$$

is a unitary operator. In the cyclic case, the unitary operator becomes a complex number of modulus one. Using the operator notation $\chi_{\psi^*\zeta}(\zeta)$ instead of the complex number notation $\chi_\psi(\zeta)$, we have instead of (4.8)

$$\chi_{\psi_2{}^*\zeta}(\zeta)\chi_{\psi_1{}^*\psi_2{}^*\zeta}(\psi_2^*\zeta) = \chi_{\psi_1{}^*\psi_2{}^*\zeta}(\zeta) \tag{4.12}$$

almost everywhere as the equation that the multipliers must satisfy.

The spectrum of the C^*-algebra $\mathfrak{A}$ discussed above is uncomfortably large. One of the major points of Lew's thesis was to construct a smaller C^*-algebra, that generated by the "tame" operators, which holds the $U(f)$'s in its weak closure.[28] The spectrum of the C^*-algebra turns out to correspond to the set of *generalized linear functionals* on $\mathfrak{I}$: that is, the set of all $F: \mathfrak{I} \to \mathbb{R} \cup \{\infty\}$ such that F is a (not necessarily continuous) linear functional on some subspace of $\mathfrak{I}$ and equals ∞ everywhere else. As before, one can discuss representations of the C^*-algebra in terms of measures on its spectrum, but even this spectrum is unwieldy, since its elements satisfy no continuity conditions.

A modern approach, due to *Gel'fand and Vilenkin*,[29] utilizes heavily the topology of $\mathfrak{I}$. It succeeds in discussing representations in terms of measures on $\mathfrak{I}'$, the *continuous dual* of $\mathfrak{I}$. This is a much smaller space indeed than either of the spectra above.

Therefore, we shall decide to carry out the remainder of our investigation in this formalism. Appendix A is devoted to the presentation of basic definitions and results. Here, we abstract from the treatment of the

Weyl group given by Gel'fand and Vilenkin, to show how their formalism may be applied to $\mathfrak{J} \wedge \mathfrak{K}$. The development is completely parallel to the Gel'fand spectral theory, although we are now working in a much smaller measure space.

Suppose that we have a continuous unitary representation U of $\mathfrak{J}$, in $\mathfrak{H}$, with $\Omega \in \mathfrak{H}$ cyclic for the representation. Then

$$L(f) = (\Omega, U(f)\Omega) \tag{4.13}$$

defines a positive definite functional on $\mathfrak{J}$. Furthermore, $L(0) = 1$, and $L(f)$ is continuous in f by the continuity of the representation U.

Therefore, $L(f)$ is the Fourier transform of a cylindrical measure on $\mathfrak{J}'$

$$(\Omega, U(f)\Omega) = \int_{\mathfrak{J}'} e^{i(F,f)}\, d\mu(F). \tag{4.14}$$

This equation is the analog of Eq. (4.2) in the Gel'fand spectral theory. From this point on, the analogies are self-evident. There is an isomorphism between $\mathfrak{H}$ and $L^2_\mu(\mathfrak{J}')$, defined by $\Omega \to \Omega(F) \equiv 1$ and $U(f)\Omega \to e^{i(F,f)}$ and extended by linearity. $U(f)$ is the operator of multiplication by $e^{i(F,f)}$, sometimes denoted $M_{e^{i(F,f)}}$.

Finally, consider a representation $U(f)V(\psi)$ of $\mathfrak{J} \wedge \mathfrak{K}$. Define $\psi^*: \mathfrak{J}' \to \mathfrak{J}'$ by setting $(\psi^* F, f) = (F, f \circ \psi)$. ψ^* is linear and continuous in the weak topology of $\mathfrak{J}'$.

The vector $\Omega_1 = V(\psi)\Omega$ is cyclic for U, and

$$\begin{aligned} L_1(f) &= (\Omega_1, U(f)\Omega_1) \\ &= (V(\psi)\Omega, U(f)V(\psi)\Omega) \\ &= \int_{\mathfrak{J}'} e^{i(F,f)}\, d\mu_1(F). \end{aligned} \tag{4.15}$$

But

$$\begin{aligned} (\Omega_1, U(f)\Omega_1) &= (\Omega, U(f \circ \psi^{-1})\Omega) \\ &= \int_{\mathfrak{J}'} \exp\,[i(F, f \circ \psi^{-1})]\, d\mu(F) \\ &= \int_{\mathfrak{J}'} \exp\,[i(\psi^{-1*}F, f)]\, d\mu(F) \\ &= \int_{\mathfrak{J}'} e^{i(F,f)}\, d\mu(\psi^* F). \end{aligned} \tag{4.16}$$

Thus $d\mu_1(F) \equiv d\mu(\psi^* F)$.

Also, from the inner product in (4.15),

$$L_1(f) = \int_{\mathfrak{J}'} e^{i(F,f)} |[V(\psi)\Omega](F)|^2\, d\mu(F), \tag{4.17}$$

from which we see that $d\mu_1(F) = |[V(\psi)\Omega](F)|^2\, d\mu(F)$. Consequently, the measures μ_1 and μ are equivalent; i.e., μ is quasi-invariant for $\mathfrak{K}$.

Thus the representation $U(f)V(\psi)$ becomes, in $L^2_\mu(\mathfrak{J}')$,

$$[U(f)\Psi](F) = e^{i(F,f)}\Psi(F), \tag{4.18}$$

$$[V(\psi)\Psi](F) = \chi_\psi(F)\Psi(\psi^* F)\left(\frac{d\mu(\psi^* F)}{d\mu(F)}\right)^{\frac{1}{2}}, \tag{4.19}$$

where the "multiplier" $\chi_\psi(F)$ is a complex-valued function of modulus one, depending on ψ and satisfying (for each pair ψ_1, ψ_2)

$$\chi_{\psi_2}(F)\chi_{\psi_1}(\psi_2^* F) = \chi_{\psi_1 \circ \psi_2}(F) \tag{4.20}$$

almost everywhere. Again, the choice $\chi_\psi(F) \equiv 1$ satisfies (4.20), defining a representation which may or may not be equivalent to the one with which we began.

In the noncyclic case there exists a finite or countably infinite family of measures $\{\mu_n\}$ on $\mathfrak{J}'$, with $\mathfrak{H} \cong \oplus_n L^2_{\mu_n}(\mathfrak{J}')$, and the operator $U(f)$ in $\mathfrak{H}$ corresponds in $L^2_{\mu_n}(\mathfrak{J}')$ to $M_{\exp[i(F,f)]}$. It follows that there is a single cylindrical measure μ on $\mathfrak{J}'$ and a μ-measurable field of Hilbert spaces $\mathfrak{H}_F$ on $\mathfrak{J}'$, such that

$$\mathfrak{H} \cong \int_{\mathfrak{J}'}^{\oplus} \mathfrak{H}_F\, d\mu(F), \tag{4.21}$$

with $U(f)$ corresponding to the operator of multiplication by $e^{i(F,f)}$.[29] The measure μ is equivalent to the transformed measure μ_1 defined by $d\mu_1(f) = d\mu(\psi^* F)$, and $\dim \mathfrak{H}_F = \dim \mathfrak{H}_{\psi^* F}$ almost everywhere, for all $\psi \in \mathfrak{K}$.

In this noncyclic case, $V(\psi)$ becomes

$$\langle (V(\psi)\Psi)(F) \rangle = \left\langle \chi_{\psi^* F}(F)\Psi(\psi^* F)\left(\frac{d\mu(\psi^* F)}{d\mu(F)}\right)^{\frac{1}{2}} \right\rangle, \tag{4.22}$$

where $\chi_{\psi^* F}(F): \mathfrak{H}_{\psi^* F} \to \mathfrak{H}_F$ is a unitary operator. The $\chi_{\psi^* F}(F)$ satisfy

$$\chi_{\psi_2^* F}(F)\chi_{\psi_1^* \psi_2^* F}(\psi_2^* F) = \chi_{\psi_1^* \psi_2^* F}(F) \tag{4.23}$$

almost everywhere, which is the analog of Eq. (4.12).

It will be noted that the set X_{ψ_1, ψ_2} of measure zero on which (4.23) fails may depend on ψ_1 and ψ_2 in such a way that the union of a family of such sets is of nonzero measure. Some of our remarks below depend upon our ability to find a set of measure zero, *independent* of ψ_1 and ψ_2, outside of which (4.23) holds. Techniques exist in somewhat different contexts for proving such a result.[30,31] However, these techniques are not immediately applicable to the "multipliers" here; for example, since our groups are not locally compact, Haar measure cannot be defined. Gel'fand and Vilenkin do not discuss this difficulty for the Weyl group, and the conjecture that, in general, such a set exists remains to be proven.

B. Discussion

We have seen in Sec. 4A that the Gel'fand–Vilenkin formalism is the most suitable for the representation theory of a group such as $\mathfrak{J} \wedge \mathfrak{K}$. The expectation functional $L(f)$ defines fully the representation of $\mathfrak{J}$ and defines the representation of $\mathfrak{K}$ up to a phase "multiplier." Whenever possible, we may choose the cyclic vector Ω used in defining $L(f)$ to be the state of lowest energy; thus $L(f)$, when coupled with the equation of continuity, can also contain important information about the dynamics.

In this section, we impose the requirement of irreducibility on a representation of $\mathfrak{J} \wedge \mathfrak{K}$. We also examine the possibility of inequivalent representations of $\mathfrak{J} \wedge \mathfrak{K}$ arising from the "multipliers" discussed above.

Let (4.18) and (4.19) define an irreducible representation of $\mathfrak{J} \wedge \mathfrak{K}$. For $F \in \mathfrak{J}'$, we can consider the *orbit* of F under the action of $\mathfrak{K}$, that is, the set $\Delta = \{\psi^* F \,|\, \psi \in \mathfrak{K}\}$. $\mathfrak{J}'$ is clearly the union of an uncountable family of mutually disjoint orbits.

An *invariant set* (for $\mathfrak{K}$) in $\mathfrak{J}'$ is a set X such that, for all $\psi \in \mathfrak{K}$, $\psi^* X = X$. Any union of orbits is certainly an invariant set; conversely, an invariant set X can always be expressed as the union of a family of orbits.

The cylindrical measure μ on $\mathfrak{J}'$ will be called *ergodic* (for $\mathfrak{K}$) if and only if, for any measurable invariant set X, either $\mu(X) = 0$ or $\mu(\mathfrak{J}' - X) = 0$.

Theorem 4: The representation (4.18) and (4.19) of $\mathfrak{J} \wedge \mathfrak{K}$ is irreducible if and only if μ is ergodic (for $\mathfrak{K}$).

Proof: Suppose first that μ is not ergodic (for $\mathfrak{K}$). Let X be a measurable invariant set with $\mu(X) > 0$ and $\mu(\mathfrak{J}' - X) > 0$. Then $\mathcal{M} = \{\Psi \in L^2_\mu(\mathfrak{J}') \,|\, \Psi(F) = 0, \forall\, F \in X\}$ is a proper nontrivial closed subspace of $L^2_\mu(\mathfrak{J}')$. Furthermore, $U(f)\mathcal{M} \subseteq \mathcal{M}$, and $V(\psi)\mathcal{M} \subseteq \mathcal{M}$ since X is invariant (for $\mathfrak{K}$). Thus $U(f)V(\psi)$ is not an irreducible representation.

Conversely, suppose that μ is ergodic, and let $\mathcal{M} \subseteq L^2_\mu(\mathfrak{J}')$ be a closed invariant subspace of $L^2_\mu(\mathfrak{J}')$. Then $P_\mathcal{M}$, the orthogonal projection onto $\mathcal{M}$, commutes with all of the $U(f)$ and $V(\psi)$. Since P commutes with all of the $U(f)$, it is a multiplication operator in $L^2_\mu(\mathfrak{J}')$; $P_\mathcal{M} = M_{\alpha(F)}$, where $\alpha(F) = (P_\mathcal{M}\Omega)(F)$. Since $P_\mathcal{M}$ is a projection, $P^2_\mathcal{M} = P_\mathcal{M}$ and $\alpha(F) = 1$ or 0 almost everywhere. Let X be $\{F \,|\, \alpha(F) = 1\}$. If $\mu(X) = 0$, $\alpha(F) = 0$ almost everywhere, $P_\mathcal{M} = 0$, and $\mathcal{M} = \{0\}$. While, if $\mu(\mathfrak{J}' - X) = 0$, $\alpha(F) = 1$ almost everywhere, $P_\mathcal{M} = I$, and $\mathcal{M} = L^2_\mu(\mathfrak{J}')$. Since these are the only two possibilities, the representation is irreducible. QED

Let us assume that the orbits we shall be considering are measurable. Then there are two ways in which a measure μ can be ergodic: It can be concentrated on a single orbit, or else every orbit can be of μ-measure zero. In Ref. 5, examples of both situations are discussed in more general context.[32]

Next we pose the question of whether (and under what circumstances) different systems of "multipliers" lead to unitarily inequivalent representations. We shall make the explicit assumption that Eq. (4.23) holds for *all* ψ_1 and ψ_2 outside of a fixed set X of measure zero.

Let us write

$$\mathfrak{J}' = \bigcup_{\gamma \in \Gamma} \Delta_\gamma,$$

where the Δ_γ are orbits, and suppose that F_0^γ is a fixed element of Δ_γ for each $\gamma \in \Gamma$. Unless $\Delta_\gamma \subseteq X$, choose $F_0^\gamma \notin X$. For any point $F \in \Delta_\gamma$, there exists $\psi \in \mathfrak{K}$ such that $F = \psi^* F_0^\gamma$; ψ need not be unique. With each point $F \in \mathfrak{J}'$, we associate the Hilbert space $\mathfrak{K}_F$, which is a replica of the complex numbers. If we are given the unitary operators $\chi_F(F_0^\gamma): \mathfrak{K}_F \to \mathfrak{K}_{F_0^\gamma}$ for all $F \in \Delta_\gamma$, $\Delta_\gamma \nsubseteq X$, then we can reconstruct $\chi_F(G)$ for all F, $G \in \Delta_\gamma$, using Eq. (4.23). In fact, $\chi_F(G) = \chi_G(F_0^\gamma)^{-1}\chi_F(F_0^\gamma)$. It is clear that the $\chi_F(F_0^\gamma)$ can be specified independently of any algebraic condition.

Theorem 5: Let $\chi_F(F_0^\gamma)$ and $\chi'_F(F_0^\gamma)$, respectively, be defined as above by two systems of "multipliers," each satisfying (4.23) outside of a fixed set of measure zero.

Then χ and χ' define unitarily equivalent representations in the Gel'fand–Vilenkin formalism if and only if the function $\beta(F) = \alpha(\gamma)\chi_F(F_0^\gamma)^{-1}\chi'_F(F_0^\gamma)$ is a measurable function on $\mathfrak{J}'$, for some function $\alpha(\gamma)$ with modulus one.

Proof: If $\beta(F)$ is measurable, then it is easily verifiable from (4.18) and (4.19) that the operation $M_{\beta(F)}$ of multiplication by $\beta(F)$ defines a unitary equivalence between the two representations.

Conversely, if Q is a unitary equivalence, $Q = M_{\sigma(F)}$ for some measurable function σ, with $|\sigma(F)| = 1$ almost everywhere.

Then

$$\sigma(F)\chi'_{\psi^* F}(F)\overline{\sigma(\psi^* F)} = \chi_{\psi^* F}(F)$$

almost everywhere, whence $\sigma(F_0^\gamma)\chi'_F(F_0^\gamma)\chi_F(F_0^\gamma)^{-1} = \sigma(F)$, which is measurable. QED

It may be remarked that the introduction of "multipliers" which are not identically one, in the action of $V(\psi)$, corresponds to the addition of real multiplication operators (functions of ρ) to the currents $J(\mathbf{g})$.

C. The Fock Representation

Let us consider the 1-particle Fock representation, defined in $L^2(\mathbb{R}^s)$ by the equations

$$U(f)\Psi(\mathbf{x}) = e^{if(\mathbf{x})}\Psi(\mathbf{x}), \tag{4.24}$$

$$V(\psi)\Psi(\mathbf{x}) = \Psi(\psi(\mathbf{x}))\left[\det\left(\frac{\partial\psi^k}{\partial x^l}(\mathbf{x})\right)\right]^{\frac{1}{2}}. \tag{4.25}$$

Take the cyclic vector to be $\Omega(\mathbf{x}) = \pi^{-\frac{1}{4}s}\exp(-\frac{1}{2}|\mathbf{x}|^2)$. Then

$$(\Omega, U(f)\Omega) = \pi^{-\frac{1}{2}s}\int_{\mathbb{R}^s} e^{if(\mathbf{x})}\exp(-|\mathbf{x}|^2)\,d\mathbf{x}. \tag{4.26}$$

Comparing (4.26) with (4.14), we conclude that μ is concentrated on the family of evaluation functionals (or δ functions) $F_{\mathbf{x}}$ defined by $(F_{\mathbf{x}}, f) = f(\mathbf{x})$, with $d\mu(F_{\mathbf{x}}) = \pi^{-\frac{1}{2}s}\exp(-|\mathbf{x}|^2)\,d\mathbf{x}$.

Theorem 6: We shall verify directly that the set $\{F_{\mathbf{x}} \mid \mathbf{x} \in R^s\}$ is a measurable set in $\mathfrak{J}'$; i.e., it can be obtained from cylinder sets by countable processes.

Proof: Let $f_0(\mathbf{x}) = \exp(-\frac{1}{2}|\mathbf{x}|^2)$ and $f_j(\mathbf{x}) = x_j\exp(-\frac{1}{2}|\mathbf{x}|^2)$ be elements of $\mathfrak{J}$, for $j = 1, \cdots, s$. Let $f_{j+1}, f_{j+2}, \cdots$ be a countable dense set in $\mathfrak{J}$. Let $A_k = \{(f_0(\mathbf{x}), \cdots, f_k(\mathbf{x})) \in \mathbb{R}^{k+1} \mid \mathbf{x} \in \mathbb{R}^s\}$ and

$$Q_k = \{F \in \mathfrak{J}' \mid ((F, f_0), \cdots, (F, f_k)) \in A_k\}.$$

Then $Q = \bigcap_{k=0}^{\infty} Q_k$ coincides with $\{F_{\mathbf{x}} \mid \mathbf{x} \in \mathbb{R}^s\}$, as follows. Clearly, $F_{\mathbf{x}} \in Q_k$ for all k, and, if $F \in Q_s$, then $((F, f_0), \cdots, (F, f_s)) = (f_0(\mathbf{x}), \cdots, f_s(\mathbf{x}))$ for some $\mathbf{x}$. But, by the definition of $f_0, \cdots, f_s$, the $(s+1)$-tuple $(f_0(\mathbf{x}), \cdots, f_s(\mathbf{x}))$ determines $\mathbf{x}$ uniquely. Thus, $F \in Q_{s+m}$ implies

$$((F, f_0), \cdots, (F, f_{s+m})) = (f_0(\mathbf{x}), \cdots, f_{s+m}(\mathbf{x})),$$

where $\mathbf{x}$ is independent of m, and, if $F \in Q$, $F = F_{\mathbf{x}}$ on a dense set and hence everywhere on $\mathfrak{J}$ by continuity. QED

The measurable set $\{F_{\mathbf{x}} \mid \mathbf{x} \in \mathbb{R}^s\}$ is a single orbit in $\mathfrak{J}'$ under the action of $\mathfrak{K}$, since, in fact, $\psi^* F_{\mathbf{x}} = F_{\psi(\mathbf{x})}$ for $\psi \in \mathfrak{K}$.

The measure associated with the 2-particle Fock representation is concentrated on the orbit

$$\{F_{\mathbf{x}_1} + F_{\mathbf{x}_2} \mid \mathbf{x}_1 \neq \mathbf{x}_2\}$$

in $\mathfrak{J}'$. Similar orbits can be constructed for all of the n-particle representations.

Next, we show how two inequivalent representations can be constructed with the same measure on the same orbit, using different systems of "multipliers."

Let $\mathbf{x} \in \mathbb{R}^3$ and

$$\sigma(\mathbf{x}) = (x^1 + ix^2)/[(x^1)^2 + (x^2)^2]^{\frac{1}{2}} \tag{4.27}$$

for x^1 or $x^2 \neq 0$. Then $|\sigma(\mathbf{x})|^2 = 1$, and $\sigma(-\mathbf{x}) = -\sigma(\mathbf{x})$. In the 2-particle, 3-dimensional Fock representation, consider the unitary operator $Q: \mathfrak{K}_s \to \mathfrak{K}_a$ given by $Q\Psi(\mathbf{x}_1, \mathbf{x}_2) = \sigma(\mathbf{x}_1 - \mathbf{x}_2)\Psi(\mathbf{x}_1, \mathbf{x}_2)$. Then $QU_s(f) = U_a(f)Q$, but, as we have seen in Theorem 3, we cannot have $QV_s(\psi) = V_a(\psi)Q$. In fact,

$$QV_s(\psi) = [\sigma(\mathbf{x}_1 - \mathbf{x}_2)/\sigma(\psi(\mathbf{x}_1) - \psi(\mathbf{x}_2))]V_a(\psi)Q. \tag{4.28}$$

Let us look at these representations in the Gel'fand–Vilenkin formalism. $\mathfrak{K}_s$ is isomorphic to $L^2_\mu(\{F_{\mathbf{x}_1} + F_{\mathbf{x}_2}\})$ by the identification

$$\Psi(\mathbf{x}_1, \mathbf{x}_2) = \pi^{-\frac{3}{2}}\exp[-\tfrac{1}{2}(\mathbf{x}_1^2 + \mathbf{x}_2^2)]\Psi(F_{\mathbf{x}_1} + F_{\mathbf{x}_2}).$$

$U_s(f)$ and $V_s(\psi)$ act in $L^2_\mu(\{F_{\mathbf{x}_1} + F_{\mathbf{x}_2}\})$ by the equations

$$U_s(f)\Psi(F_{\mathbf{x}_1} + F_{\mathbf{x}_2}) = \exp\{i[f(\mathbf{x}_1) + f(\mathbf{x}_2)]\}\Psi(F_{\mathbf{x}_1} + F_{\mathbf{x}_2}), \tag{4.29}$$

$$V_s(\psi)\Psi(F_{\mathbf{x}_1} + F_{\mathbf{x}_2}) = \Psi(F_{\psi(\mathbf{x}_1)} + F_{\psi(\mathbf{x}_2)}) \times \prod_{k=1}^{2}\left[\frac{\exp[-|\psi(\mathbf{x}_k)|^2]}{\exp(-|\mathbf{x}_k|^2)}\det\left(\frac{\partial\psi^l}{\partial x^j}(\mathbf{x}_k)\right)\right]^{\frac{1}{2}}. \tag{4.30}$$

$\mathfrak{K}_a$ is isomorphic to $L^2_\mu(\{F_{\mathbf{x}_1} + F_{\mathbf{x}_2}\})$ by the identification

$$\Psi(\mathbf{x}_1, \mathbf{x}_2) = \sigma(\mathbf{x}_1 - \mathbf{x}_2)\pi^{-\frac{3}{2}} \times \exp[-\tfrac{1}{2}(\mathbf{x}_1^2 + \mathbf{x}_2^2)]\Psi(F_{\mathbf{x}_1} + F_{\mathbf{x}_2}).$$

Then $U_a(f) = U_s(f)$ in $L^2_\mu(\{F_{\mathbf{x}_1} + F_{\mathbf{x}_2}\})$, while

$$V_a(\psi) = [\sigma(\psi(\mathbf{x}_1) - \psi(\mathbf{x}_2))/\sigma(\mathbf{x}_1 - \mathbf{x}_2)]V_s(\psi).$$

In short, V_a corresponds to the system of "multipliers"

$$\chi_\psi(F_{\mathbf{x}_1} + F_{\mathbf{x}_2}) = \sigma(\psi(\mathbf{x}_1) - \psi(\mathbf{x}_2))/\sigma(\mathbf{x}_1 - \mathbf{x}_2), \tag{4.31}$$

while the "multipliers" for V_s are identically one.

Thus, *n-particle representations having different exchange symmetries correspond in the Gel'fand–Vilenkin formalism to representations having different systems of "multipliers."*

With

$$J_a(\mathbf{g}) = (2i)^{-1}\sum_{k=1}^{2}[\mathbf{g}(\mathbf{x}_k)\cdot\nabla_k + \nabla_k\cdot\mathbf{g}(\mathbf{x}_k)] \quad \text{in } \mathfrak{K}_a, \tag{4.32}$$

we may define $J'(\mathbf{g})$ to satisfy $QJ'(\mathbf{g}) = J_a(\mathbf{g})Q$. Then,

using (4.27), we obtain

$$J'(\mathbf{g}) = J_s(\mathbf{g}) - \left(\frac{(x_1^2 - x_2^2)[g^1(\mathbf{x}_1) - g^1(\mathbf{x}_2)] - (x_1^1 - x_2^1)[g^2(\mathbf{x}_1) - g^2(\mathbf{x}_2)]}{(x_1^1 - x_2^1)^2 + (x_1^2 - x_2^2)^2}\right), \tag{4.33}$$

illustrating our remark that the introduction of "multipliers" which are not identically one corresponds to the addition of real multiplication operators to the currents $J(\mathbf{g})$.[33] $J'(\mathbf{g})$ corresponds to a "fermion" representation of the current algebra, acting in $\mathcal{H}_s$.

One may verify directly that $J'(\mathbf{g})$ satisfies the current algebra (1.10)–(1.12).

D. The Theory in a Box[33]

Grodnik and Sharp have studied the algebra obtained by setting

$$\rho_{\mathbf{m}} = \int e^{-i\mathbf{m}\cdot\mathbf{x}}\rho(\mathbf{x})\,d\mathbf{x} \tag{4.34}$$

and

$$J_{\mathbf{n}}^k = \int e^{-i\mathbf{n}\cdot\mathbf{x}}J^k(\mathbf{x})\,d\mathbf{x}, \tag{4.35}$$

where ρ and J satisfy (1.5)–(1.7). It follows that

$$[\rho_{\mathbf{m}}, \rho_{\mathbf{n}}] = 0, \tag{4.36}$$

$$[\rho_{\mathbf{m}}, J_{\mathbf{n}}^k] = m_k\rho_{\mathbf{m}+\mathbf{n}}, \tag{4.37}$$

$$[J_{\mathbf{m}}^j, J_{\mathbf{n}}^k] = m_kJ_{\mathbf{m}+\mathbf{n}}^j - n_jJ_{\mathbf{m}+\mathbf{n}}^k. \tag{4.38}$$

A formal representation of the $(\rho_{\mathbf{m}}, J_{\mathbf{n}})$-algebra is obtained by setting

$$\rho_{\mathbf{m}} = z_{\mathbf{m}}, \tag{4.39}$$

$$J_{\mathbf{n}}^k = \sum_{\ell=-\infty}^{\infty} \ell_k z_{\mathbf{n}+\ell}\frac{\partial}{\partial z_\ell}, \tag{4.40}$$

where we imagine these operators to act on some space of functions of infinitely many complex variables $z_{\mathbf{m}}$.

The purpose of this subsection is to put the theory in a box on a rigorous footing by using some of the representation theory which we have developed. In particular, we wish to identify the "space of functions of infinitely many complex variables."

Let us take the "box" to be the torus T^s, an s cube of length $2a$ in each spatial direction, with points on opposite boundaries of the s cube identified. The choice of the torus imposes periodic boundary conditions on the testing functions. Our testing-function space will be $C_\infty(T^s)$, the space of real-valued, infinitely differentiable functions on the torus.

A topology may be defined on $C_\infty(T^s)$ by means of the infinite set of norms

$$\|f\|_n = \max_{|\mathbf{m}|\le n}\sup_{\mathbf{x}\in T^s}\left|\frac{\partial^{|\mathbf{m}|}}{(\partial x^1)^{m_1}\cdots(\partial x^s)^{m_s}}f(\mathbf{x})\right|, \tag{4.41}$$

for $n = 0, 1, 2, \cdots$. Under this topology, $C_\infty(T^s)$ becomes a nuclear space in the sense of Gel'fand and Vilenkin.

Let $\mathfrak{W}$ be the group of all C_∞ diffeomorphisms $\psi: T^s \to T^s$ under composition. A "theory in a box" is given by a strongly continuous unitary representation of the semidirect product $C_\infty(T^s) \wedge \mathfrak{W}$, where the semidirect product law is defined by $(f, \psi) \to f \circ \psi$.[34]

The 1-particle Fock representation is given by

$$U(f)\Psi(\mathbf{x}) = e^{if(\mathbf{x})}\Psi(\mathbf{x}), \tag{4.42}$$

$$V(\psi)\Psi(\mathbf{x}) = \Psi(\psi(\mathbf{x}))\left[\det\left(\frac{\partial\psi^k}{\partial x^j}(\mathbf{x})\right)\right]^{\frac{1}{2}} \tag{4.43}$$

on $L^2(T^s)$. $\Omega(\mathbf{x}) \equiv [(2a)^s]^{-\frac{1}{2}}$ defines a cyclic vector. Derivatives are always taken with respect to the system of local coordinates on the torus T^s implied by its identification with an s cube.

It should be mentioned that, for the theory in a box, the unitary equivalence between n-particle representations having different exchange symmetries breaks down even in the case of one spatial dimension (compare Theorems 1 and 3). Roughly speaking, this is because a flow may be found on the torus which exchanges any two points, while such a flow does not exist on the real line.

The next two theorems are devoted to expressing the continuous dual of $C_\infty(T^s)$ as a space of sequences of complex numbers.

Consider the Fourier functions on T^s,

$$f_{(\mathbf{m})}(\mathbf{x}) = [(2a)^s]^{-\frac{1}{2}}e^{-i\mathbf{m}\cdot\mathbf{x}\pi/a}. \tag{4.44}$$

It is well known that if $f \in L^2(T^s)$, we can expand

$$f = \sum_{\mathbf{m}} c_{\mathbf{m}}f_{(\mathbf{m})},$$

where the sum is over $m_1, \cdots, m_s = -\infty, \cdots, \infty$ and the series converges in $L^2(T^s)$. Furthermore,

$$c_{\mathbf{m}} = [(2a)^s]^{-\frac{1}{2}}\int e^{i\mathbf{m}\cdot\mathbf{x}\pi/a}f(\mathbf{x})\,d\mathbf{x}. \tag{4.45}$$

If we identify the element f in $L^2(T^s)$ with the indexed set $\{c_{\mathbf{m}}\}$ of its Fourier coefficients, the requirement that f be square integrable imposes the requirement that

$$\sum_{\mathbf{m}} |c_{\mathbf{m}}|^2 < \infty.$$

Theorem 7: Let f be a square-integrable function on T^s, and let $\{c_{\mathbf{m}}\}$ be the indexed set of its Fourier coefficients. Then $f \in C_\infty(T^s)$ if and only if, for any polynomial P in $m_1, \cdots, m_s$,

$$\sum_{\mathbf{m}} |P(\mathbf{m})|^2 \, |c_{\mathbf{m}}|^2 < \infty.$$

We say then that $c_{\mathbf{m}} \to 0$ rapidly as $\mathbf{m} \to \infty$.

Proof: It suffices to show that, as an element of the Hilbert space $L^2(T^s)$, f is in the domain of $A = -i(d/dx_1)$ if and only if the series $\sum_{\mathbf{m}} m_1 c_{\mathbf{m}} f_{(\mathbf{m})}$ converges in $L^2(T^s)$. But A is a self-adjoint operator in $L^2(T^s)$. If $\sum_{\mathbf{m}} m_1 c_{\mathbf{m}} f_{(\mathbf{m})}$ converges, then f is in the domain of A by virtue of the fact that A is closed.

Conversely, suppose f to be in the domain of A, and let $Af = \sum_{\mathbf{m}} d_{\mathbf{m}} f_{(\mathbf{m})}$. With $|\mathbf{m}| = \sum_{k=1}^{s} m_k$,

$$\begin{aligned} A \sum_{|\mathbf{m}|>M} c_{\mathbf{m}} f_{(\mathbf{m})} &= A\left(f - \sum_{|\mathbf{m}|\le M} c_{\mathbf{m}} f_{(\mathbf{m})}\right) \\ &= \sum_{|\mathbf{m}|\le M} [d_{\mathbf{m}} - (-\pi/a) m_1 c_{\mathbf{m}}] f_{(\mathbf{m})} \\ &\quad + \sum_{|\mathbf{m}|>M} d_{\mathbf{m}} f_{(\mathbf{m})}. \end{aligned} \tag{4.46}$$

But, if $|\mathbf{n}| \le M$,

$$\begin{aligned} &\left(A \sum_{|\mathbf{m}|>M} c_{\mathbf{m}} f_{(\mathbf{m})}, f_{(\mathbf{n})}\right) \\ &\quad = \left(\sum_{|\mathbf{m}|>M} c_{\mathbf{m}} f_{(\mathbf{m})}, (-\pi/a) n_1 f_{(\mathbf{n})}\right) = 0, \end{aligned} \tag{4.47}$$

and in (4.46), for $|\mathbf{m}| \le M$, $d_{\mathbf{m}} = (-\pi/a) m_1 c_{\mathbf{m}}$. As M is arbitrary, the series $\sum_{\mathbf{m}} m_1 c_{\mathbf{m}} f_{(\mathbf{m})}$ converges. QED

Requiring f to be real valued corresponds to the auxiliary condition $\bar{c}_{\mathbf{m}} = c_{-\mathbf{m}}$ on the indexed set $\{c_{\mathbf{m}}\}$. Then we shall write $C = \{c_{\mathbf{m}}\}$.

Theorem 8: Let F be an element of the continuous dual of $C_\infty(T^s)$, and set $\lambda_{\mathbf{m}} = (F, f_{(\mathbf{m})})$. Then $\lambda_{\mathbf{m}}$ is of polynomial growth in $\mathbf{m}$; i.e., there exists a polynomial $P(\mathbf{m})$ with $|\lambda_{\mathbf{m}}| \le |P(\mathbf{m})|$ for all $\mathbf{m}$. An indexed set $\{\lambda_{\mathbf{m}}\}$ of complex numbers, with $\lambda_{\mathbf{m}}$ of polynomial growth in $\mathbf{m}$ and with the reality condition $\bar{\lambda}_{\mathbf{m}} = \lambda_{-\mathbf{m}}$, will be denoted Λ. Any such Λ defines a continuous linear functional on $C_\infty(T^s)$.

Proof: With $f = \sum_{\mathbf{m}} c_{\mathbf{m}} f_{(\mathbf{m})}$, we have

$$(F, f) = \sum_{\mathbf{m}} \lambda_{\mathbf{m}} c_{\mathbf{m}}.$$

In order for this sum to converge for all C, it is clearly necessary and sufficient that $\lambda_{\mathbf{m}}$ be of polynomial growth in $\mathbf{m}$. We need only show that the linear functional on $C_\infty(T^s)$ defined by Λ is continuous in the topology given by (4.41).

Suppose that for all n, $\lim \|f_k\|_n = 0$ as $k \to \infty$, for $f_k \in C_\infty(T^s)$. Let $f_k = \sum_{\mathbf{m}} c_{\mathbf{m}}^k f_{(\mathbf{m})}$. We shall show that $\lim (F, f_k) = 0$ as $k \to \infty$, where (F, f_k) is defined from Λ by the formula

$$(F, f_k) = \sum_{\mathbf{m}} \lambda_{\mathbf{m}} c_{\mathbf{m}}^k.$$

It is not hard to demonstrate that the convergence of $\|f_k\|_n$ to zero, for all n, implies that

$$\lim_{k\to\infty} \sup_{\mathbf{x}\in T^s} \left| \sum_{\mathbf{m}} Q(\mathbf{m}) c_{\mathbf{m}}^k f_{(\mathbf{m})}(\mathbf{x}) \right|^2 = 0$$

for all polynomials $Q(\mathbf{m})$, whence

$$\lim_{k\to\infty} \sum_{\mathbf{m}} |Q(\mathbf{m}) c_{\mathbf{m}}^k|^2 = 0 \quad \text{and} \quad \lim_{k\to\infty} \sup_{\mathbf{m}} |Q(\mathbf{m}) c_{\mathbf{m}}^k| = 0.$$

But then

$$\begin{aligned} \left| \sum_{\mathbf{m}} \lambda_{\mathbf{m}} c_{\mathbf{m}}^k \right| &\le \sum_{\mathbf{m}} \left| \frac{\lambda_{\mathbf{m}}}{Q(\mathbf{m})} \right| \cdot |Q(\mathbf{m}) c_{\mathbf{m}}^k| \\ &\le \sup_{\mathbf{m}} |Q(\mathbf{m}) c_{\mathbf{m}}^k| \cdot \sum_{\mathbf{m}} \left| \frac{\lambda_{\mathbf{m}}}{Q(\mathbf{m})} \right|, \end{aligned}$$

which approaches zero as k approaches infinity, Q being a polynomial of sufficiently high degree which bounds $|\lambda_{\mathbf{m}}|$. QED

Now we can apply the representation theory of Gel'fand and Vilenkin. Let us first redefine

$$\rho_{\mathbf{m}} = \rho(f_{(\mathbf{m})}) = [(2a)^s]^{-\frac{1}{2}} \int e^{-i\mathbf{m}\cdot\mathbf{x}\pi/a} \rho(\mathbf{x}) \, d\mathbf{x}, \tag{4.48}$$

$$J_{\mathbf{n}}^k = J^k(f_{(\mathbf{n})}) = [(2a)^s]^{-\frac{1}{2}} \int e^{-i\mathbf{n}\cdot\mathbf{x}\pi/a} J^k(\mathbf{x}) \, d\mathbf{x}, \tag{4.49}$$

whence the current algebra (4.36)–(4.38) becomes

$$[\rho_{\mathbf{m}}, \rho_{\mathbf{n}}] = 0, \tag{4.50}$$

$$[\rho_{\mathbf{m}}, J_{\mathbf{n}}^k] = (\pi/a)[(2a)^s]^{-\frac{1}{2}} m^k \rho_{\mathbf{m}+\mathbf{n}}, \tag{4.51}$$

$$[J_{\mathbf{m}}^j, J_{\mathbf{n}}^k] = (\pi/a)[(2a)^s]^{-\frac{1}{2}} (m^k J_{\mathbf{m}+\mathbf{n}}^j - n^j J_{\mathbf{m}+\mathbf{n}}^k). \tag{4.52}$$

Theorem 9: Let U be a (strongly) continuous, cyclic, unitary representation of $C_\infty(T^s)$ with cyclic vector Ω. Then we can realize $\rho_{\mathbf{m}}$ in such a representation as multiplication by $\lambda_{\mathbf{m}}$, on the space of functions $\Psi(\Lambda)$ square integrable with respect to a cylindrical measure μ on the dual of $C_\infty(T^s)$. The cyclic vector may be realized as $\Omega(\Lambda) \equiv 1$. The domain

$$\left\{\Psi(\Lambda) = \sum_{k=1}^{n} a_k \exp\,[i(\Lambda, C^k)]\right\}$$

is dense in $L_\mu^2(C_\infty(T^s)')$, where $(\Lambda, C) = \sum_{\mathbf{m}} \lambda_{\mathbf{m}} c_{\mathbf{m}}$.

Proof: By (4.14), we can realize $\mathcal{K}$ as $L^2_\mu(C_\infty(T^s)')$, with $\Omega(\Lambda) \equiv 1$, and $U(C)\Psi(\Lambda) = e^{i(\Lambda,C)}\Psi(\Lambda)$, where μ is a cylindrical measure in $C_\infty(T^s)'$. In such a representation, the continuous 1-parameter groups

$$U(\tfrac{1}{2}t(f_{(\mathbf{m})} + f_{(-\mathbf{m})})) \quad \text{and} \quad U((t/2i)(f_{(\mathbf{m})} - f_{(-\mathbf{m})}))$$

act on $\Psi(\Lambda)$ by

$$U(\cdots, 0, c_{-\mathbf{m}} = \tfrac{1}{2}t, \cdots 0 \cdots, c_{\mathbf{m}} = \tfrac{1}{2}t, 0, \cdots)\Psi(\Lambda) = \exp\,[it\tfrac{1}{2}(\lambda_{-\mathbf{m}} + \lambda_{\mathbf{m}})]\Psi(\Lambda) \quad (4.53)$$

and

$$U(\cdots, 0, c_{-m} = -t/2i, \cdots, 0, \cdots, c_{\mathbf{m}} = t/2i, 0, \cdots)\Psi(\Lambda) = \exp\,\{it[(2i)^{-1}(-\lambda_{-\mathbf{m}} + \lambda_{\mathbf{m}})]\}\Psi(\Lambda). \quad (4.54)$$

The respective infinitesimal generators are (by Stone's theorem) the self-adjoint operators

$$\rho^+_{\mathbf{m}} = \tfrac{1}{2}(\lambda_{\mathbf{m}} + \lambda_{-\mathbf{m}}) = \operatorname{Re}\lambda_{\mathbf{m}}, \quad (4.55)$$

$$\rho^-_{\mathbf{m}} = \tfrac{1}{2}(\lambda_{\mathbf{m}} - \lambda_{-\mathbf{m}}) = \operatorname{Im}\lambda_{\mathbf{m}}. \quad (4.56)$$

On the common domain of $\rho^+_{\mathbf{m}}$ and $\rho^-_{\mathbf{m}}$, we immediately recover $\rho_{\mathbf{m}} = \rho^+_{\mathbf{m}} + i\rho^-_{\mathbf{m}}$ and $\rho_{-\mathbf{m}} = \rho^+_{\mathbf{m}} - i\rho^-_{\mathbf{m}}$. QED

Theorem 9 says that the formal representation $\rho_{\mathbf{m}} = \lambda_{\mathbf{m}}$ is actually a perfectly general representation of the $\rho_{\mathbf{m}}$. The specific choice is made only when a cylindrical measure is chosen on $\{\Lambda\}$.

In Sec. 5, we shall develop sufficient conditions on the cylindrical measure μ for the simultaneous recovery of all the $\rho_{\mathbf{m}}$ and $J_{\mathbf{n}}$ on a common dense invariant domain, in a representation of $C_\infty(T^s) \wedge \mathcal{W}$. The domain D turns out to include the set

$$D' = \{\Psi(\Lambda) = \beta(\operatorname{Re}\lambda_{\mathbf{m}_1}, \operatorname{Im}\lambda_{\mathbf{m}_1}, \cdots, \operatorname{Re}\lambda_{\mathbf{m}_n}, \operatorname{Im}\lambda_{\mathbf{m}_n}); (\forall n)(\forall \beta \in \mathcal{O}_M(2n))(\forall \mathbf{m}_1, \cdots, \mathbf{m}_n)\},$$

where $\mathcal{O}_M(n)$ is the class of complex-valued C_∞ functions of n real variables which, together with their derivatives of all orders, are of polynomial growth at ∞. D' is a domain of essential self-adjointness for $\rho^+_{\mathbf{m}}$ and $\rho^-_{\mathbf{m}}$, for all $\mathbf{m}$.

Under the conditions of Sec. 5, the operator $J^k_{\mathbf{m}}$ can be represented on the domain D' in $L^2_\mu(\{\Lambda\})$ by

$$J^k_{\mathbf{m}} = -\frac{\pi}{a}\frac{1}{[(2a)^s]^{\frac{1}{2}}}\sum_{\ell=-\infty}^{\infty}\ell^k\lambda_{\mathbf{m}+\ell}\frac{\partial}{\partial\lambda_\ell} + M_{(J_{\mathbf{m}}{}^k\Omega)(\Lambda)}, \quad (4.57)$$

where M denotes the operation of pointwise multiplication.[5] In particular, the sum

$$\sum_{\ell=-\infty}^{\infty}\ell^k\lambda_{\mathbf{m}+\ell}\frac{\partial}{\partial\lambda_\ell} \quad (4.58)$$

is a well-defined operator on D', whereas the individual derivative $\partial/\partial\lambda_\ell$ may not be well defined.[23]

Thus the representation (4.57) is also a general form; the specific choice of representation is made when a cylindrical measure quasi-invariant under the group $\mathcal{W}$ is placed on $\{\Lambda\}$.

The 1-particle Fock representation in a box corresponds to a cylindrical measure μ concentrated on the evaluation functionals in the dual of $C_\infty(T^s)$. If $F_{\mathbf{x}}$ is the evaluation functional at the point $\mathbf{x}$, then

$$(F_{\mathbf{x}}, f_{(\mathbf{m})}) = \lambda_{\mathbf{m}} = (2a)^{-\frac{1}{2}s}e^{-i\mathbf{m}\cdot\mathbf{x}\pi/a}.$$

Thus, in $\{\Lambda\}$, μ is concentrated on

$$\{\Lambda \mid \lambda_{\mathbf{m}} = (2a)^{-\frac{1}{2}s}e^{-i\mathbf{m}\cdot\mathbf{x}\pi/a};\ \mathbf{x} \in \mathbb{R}^s\}.$$

The identity

$$\lambda_{\mathbf{m}_1}\lambda_{\mathbf{m}_2} = (2a)^{-\frac{1}{2}s}\lambda_{\mathbf{m}_1+\mathbf{m}_2} \quad (4.59)$$

holds almost everywhere, from which we obtain the operator equation

$$\rho_{\mathbf{m}}\rho_{\mathbf{n}} = [(2a)^s]^{-\frac{1}{2}}\rho_{\mathbf{m}+\mathbf{n}}. \quad (4.60)$$

(4.60) may be termed the "1-particle identity." [23]

In this example, we easily see that $\partial/\partial\lambda_\ell$ is not a well-defined operator, while the sum (4.58) nevertheless is well defined.

The 2-particle Fock representation corresponds to a cylindrical measure μ concentrated on the functionals $F_{\mathbf{x}_1} + F_{\mathbf{x}_2}$ with $\mathbf{x}_1 \neq \mathbf{x}_2$, and correspondingly on

$$\{\Lambda \mid \lambda_{\mathbf{m}} = (2a)^{-\frac{1}{2}s}[e^{-i\mathbf{m}\cdot\mathbf{x}_1\pi/a} + e^{-i\mathbf{m}\cdot\mathbf{x}_2\pi/a}];\ \mathbf{x}_1 \neq \mathbf{x}_2\}.$$

Therefore the "2-particle identity"

$$\lambda_{\mathbf{m}_1}\lambda_{\mathbf{m}_2}\lambda_{\mathbf{m}_3} = (2a)^{-\frac{1}{2}s}[\lambda_{\mathbf{m}_1}\lambda_{\mathbf{m}_2+\mathbf{m}_3} + \lambda_{\mathbf{m}_2}\lambda_{\mathbf{m}_1+\mathbf{m}_3} + \lambda_{\mathbf{m}_3}\lambda_{\mathbf{m}_1+\mathbf{m}_2}] - 2(2a)^{-s}\lambda_{\mathbf{m}_1+\mathbf{m}_2+\mathbf{m}_3} \quad (4.61)$$

holds almost everywhere, or

$$\rho_{\mathbf{m}_1}\rho_{\mathbf{m}_2}\rho_{\mathbf{m}_3} = [(2a)^s]^{-\frac{1}{2}}(\rho_{\mathbf{m}_1}\rho_{\mathbf{m}_2+\mathbf{m}_3} + \rho_{\mathbf{m}_2}\rho_{\mathbf{m}_1+\mathbf{m}_3} + \rho_{\mathbf{m}_3}\rho_{\mathbf{m}_1+\mathbf{m}_2}) - 2[(2a)^s]^{-1}\rho_{\mathbf{m}_1+\mathbf{m}_2+\mathbf{m}_3}. \quad (4.62)$$

Similarly, the n-particle Fock representation leads always to an "n-particle identity." When the n-particle identity is satisfied, the higher identities are automatically satisfied, but the converse is, of course, false.

The representations in this subsection are the Fourier transforms of the so-named "functional representations" of the (ρ, J)-algebra. Functional representations of the theory in a box are studied extensively by Grodnik and Sharp.

5. RECOVERING THE CURRENT ALGEBRA

We have seen in Sec. 3 that, instead of representations of the algebra of (smeared) currents, it is sensible to study representations of the corresponding exponentiated formalism. In the present section, we develop some sufficient conditions for the recovery of the current algebra to be possible in such a representation. The current algebra is to be recovered from the infinitesimal generators of 1-parameter unitary subgroups, on a dense domain in the Hilbert space.

The sufficient conditions are expressed as properties of the cylindrical measure in the Gel'fand–Vilenkin formalism described in Sec. 4. We shall see some of the advantages of the Gel'fand–Vilenkin formalism over the other approaches mentioned in that section.

Our development will be for the group $\mathfrak{I} \wedge \mathfrak{K}$; however, it is easily generalized to include $C_\infty(T^s) \wedge \mathfrak{W}$, the Weyl group, etc.[5] Finally, we shall assume for simplicity that the "multiplier" $\chi_{\psi * F}(F) \equiv 1$.

Theorem 10: Let $U(f)V(\psi)$ be a strongly continuous unitary representation of $\mathfrak{I} \wedge \mathfrak{K}$ in a Hilbert space $\mathfrak{H}$. Let

$$\rho(f) = \lim_{t \to 0} (it)^{-1}[U(f) - I] \tag{5.1}$$

and

$$J(\mathbf{g}) = \lim_{t \to 0} (it)^{-1}[V(\varphi_t^{\mathbf{g}}) - I] \tag{5.2}$$

be defined by Stone's theorem on domains $D_{\rho(f)}$ and $D_{J(\mathbf{g})}$, respectively. Then, we have the following:

(a) Let $\Psi \in D_{\rho(f)}$; then, for all $\lambda \in \mathbb{R}$, $\Psi \in D_{\rho(\lambda f)}$ and $\rho(\lambda f)\Psi = \lambda\rho(f)\Psi$.

(b) Let $\Psi \in D_{\rho(f_1)}$, $D_{\rho(f_2)}$; then $\Psi \in D_{\rho(f_1+f_2)}$ and $\rho(f_1 + f_2)\Psi = \rho(f_1)\Psi + \rho(f_2)\Psi$.

(c) Let $\Psi \in D_{\rho(f_1)}$, $D_{\rho(f_2)}$; also let $\rho(f_2)\Psi \in D_{\rho(f_1)}$ and $\rho(f_1)\Psi \in D_{\rho(f_2)}$. Then $[\rho(f_1), \rho(f_2)]\Psi = 0$.

(d) Let $\Psi \in D_{\rho(f)}$ for all f, and let $\Psi \in D_{J(\mathbf{g}_1)}$; also let $J(\mathbf{g}_1)\Psi \in D_{\rho(f)}$ for all f, and $\rho(f)\Psi \in D_{J(\mathbf{g}_1)}$. Furthermore, suppose that $\rho(f)\Psi$ is continuous in f. Then, for all f, $[\rho(f), J(\mathbf{g}_1)]\Psi = i\rho(\mathbf{g}_1 \cdot \nabla f)\Psi$.

(e) Let Ψ satisfy the assumptions of (d) for $J(\mathbf{g}_1)$, $J(\lambda\mathbf{g}_1)$, $J(\mathbf{g}_2)$, and $J(\mathbf{g}_1 + \mathbf{g}_2)$. Then, for all f,

$$[\rho(f), J(\lambda\mathbf{g}_1)]\Psi = [\rho(f), \lambda J(\mathbf{g}_1)]\Psi,$$

and

$$[\rho(f), J(\mathbf{g}_1 + \mathbf{g}_2)]\Psi = [\rho(f), J(\mathbf{g}_1) + J(\mathbf{g}_2)]\Psi.$$

Let Ψ be a vector to which the three operators $\rho(f)$, $J(\mathbf{g}_1)$, and $J(\mathbf{g}_2)$ may be applied successively in any order for all f; also suppose that $\rho(f)$ and $J(\mathbf{g}_2 \cdot \nabla\mathbf{g}_1 - \mathbf{g}_1 \cdot \nabla\mathbf{g}_2)$ may be applied in either order to Ψ. Furthermore, let $\rho(f)\Psi$, $\rho(f)J(\mathbf{g}_1)\Psi$, $\rho(f)J(\mathbf{g}_2)\Psi$, and $\rho(f)J(\mathbf{g}_2 \cdot \nabla\mathbf{g}_1 - \mathbf{g}_1 \cdot \nabla\mathbf{g}_2)\Psi$ be continuous in f. Then, for all f,

$$[\rho(f), [J(\mathbf{g}_1), J(\mathbf{g}_2)]]\Psi = [\rho(f), iJ(\mathbf{g}_2 \cdot \nabla\mathbf{g}_1 - \mathbf{g}_1 \cdot \nabla\mathbf{g}_2)]\Psi.$$

Proof: (a), (b), and (c) follow easily from Stone's theorem. In order to prove (d), we first demonstrate that

$$\lim_{t \to 0} \frac{V(\varphi_t^{\mathbf{g}_1}) - I}{it} \rho(f)\Psi = \lim_{t \to 0} \frac{V(\varphi_t^{\mathbf{g}_1}) - I}{it} \rho(f \circ \varphi_{-t}^{\mathbf{g}_1})\Psi. \tag{5.3}$$

In fact,

$$\lim_{t \to 0} \left\| i^{-1}[V(\varphi_t^{\mathbf{g}_1}) - I]\rho\left(\frac{f \circ \varphi_{-t}^{\mathbf{g}_1} - f}{t} + \mathbf{g}_1 \cdot \nabla f\right)\Psi \right\|$$
$$\leq \lim_{t \to 0} \|V(\varphi_t^{\mathbf{g}_1}) - I\| \cdot \left\| \rho\left(\frac{f \circ \varphi_{-t}^{\mathbf{g}_1} - f}{t} + \mathbf{g}_1 \cdot \nabla f\right)\Psi \right\| = 0, \tag{5.4}$$

since $\|V(\varphi_t^{\mathbf{g}_1}) - I\|$ is bounded and since $\rho(f)\Psi$ is continuous in f. Also,

$$\lim_{t \to 0} i^{-1}[V(\varphi_t^{\mathbf{g}_1}) - I]\rho(\mathbf{g}_1 \cdot \nabla f)\Psi = 0 \tag{5.5}$$

by the strong continuity of V. Thus,

$$\lim_{t \to 0} \frac{V(\varphi_t^{\mathbf{g}_1}) - I}{it} [\rho(f \circ \varphi_{-t}^{\mathbf{g}_1}) - \rho(f)]\Psi$$
$$= \lim_{t \to 0} i^{-1}[V(\varphi_t^{\mathbf{g}_1}) - I]\rho\left(\frac{f \circ \varphi_{-t}^{\mathbf{g}_1} - f}{t}\right)\Psi = 0, \tag{5.6}$$

and (5.3) is proved.

Then we have

$$\begin{aligned} J(\mathbf{g}_1)\rho(f)\Psi &= \lim_{t \to 0} \frac{V(\varphi_t^{\mathbf{g}_1}) - I}{it} \rho(f)\Psi \\ &= \lim_{t \to 0} \frac{V(\varphi_t^{\mathbf{g}_1}) - I}{it} \rho(f \circ \varphi_{-t}^{\mathbf{g}_1})\Psi \\ &= \lim_{t \to 0} \lim_{s \to 0} \frac{V(\varphi_t^{\mathbf{g}_1}) - I}{it} \frac{U(sf \circ \varphi_{-t}^{\mathbf{g}_1}) - I}{is}\Psi \\ &= \lim_{t \to 0} \lim_{s \to 0} \left[\left(\frac{U(sf) - I}{is}\right)\left(\frac{V(\varphi_t^{\mathbf{g}_1}) - I}{it}\right) \right. \\ &\quad \left. - \frac{1}{it}\left(\frac{U(sf \circ \varphi_{-t}^{\mathbf{g}_1}) - I}{is} - \frac{U(sf) - I}{is}\right)\right] \\ &= \lim_{t \to 0} \rho(f) \frac{V(\varphi_t^{\mathbf{g}_1}) - I}{it}\Psi \\ &\quad - \lim_{t \to 0} \frac{1}{it} [\rho(f \circ \varphi_{-t}^{\mathbf{g}_1}) - \rho(f)]\Psi, \end{aligned} \tag{5.7}$$

where $V(\varphi_t^{g_1})\Psi \in D_{\rho(f)}$ since

$$\lim_{s\to 0} \frac{U(sf) - I}{is} V(\varphi_t^{g_1})\Psi = \lim_{s\to 0} V(\varphi_t^{g_1}) \frac{U(sf \circ \varphi_{-t}^{g_1}) - I}{is}\Psi, \quad (5.8)$$

and $\Psi \in D_{\rho(f\circ\varphi_{-t}^{g_1})}$.

In (5.7), the second term is just $-i\rho(\mathbf{g}_1 \cdot \nabla f)\Psi$. The first limit, which exists since all of the other terms exist, converges to $\rho(f)J(\mathbf{g}_1)\Psi$ because $\rho(f)$ is closed. Thus

$$J(\mathbf{g}_1)\rho(f)\Psi = \rho(f)J(\mathbf{g}_1)\Psi - i\rho(\mathbf{g}_1 \cdot \nabla f)\Psi, \quad (5.9)$$

which is the result (d).

The results in (e) are easy consequences of (a)–(d), by using finally the Jacobi identity. We omit the details. QED

It is interesting to note that the continuity of $\rho(f)\Psi$ in f is important in recovering the commutation relations.

A. Assumptions on μ

Let us now look systematically at the consequences of certain assumptions on the Gel'fand–Vilenkin measure μ, in a representation $U(f)V(\psi)$ of $\mathfrak{J} \wedge \mathcal{K}$. We shall use the Lebesgue dominated convergence theorem, which is included in Appendix B.

Assumption 1: For all n and for all $f_1, \cdots, f_n \in \mathfrak{J}$, the "moment"

$$\int |(F, f_1)|^2 \cdots |(F, f_n)|^2 \, d\mu(F) < \infty.$$

Theorem 11: Under Assumption 1, there exists a common dense invariant domain for all of the $\rho(f)$. The vector Ω cyclic for $U(\mathfrak{J})$ is in this domain.

Proof: Let $M(S_1, \cdots, S_n)$ be a monomial in n real variables. If $\Psi(F)$ is essentially bounded, then $M((F, f_1), \cdots, (F, f_n))\Psi(F)$ is square integrable. Let

$$\Phi_t = \{[U(tf) - I]/it\}M((F, f_1), \cdots, (F, f_n))\Psi(F), \quad (5.10)$$

and $\Phi = (F, f)M((F, f_1), \cdots, (F, f_n))\Psi(F)$. We argue that $\lim \|\Phi_t - \Phi\| = 0$ as $t \to 0$. In fact, $\Phi_t \to \Phi$ pointwise, while $|\Phi_t(F)| \leq |\chi(F)|$ almost everywhere, where

$$\chi(F) = |\Phi(F)| \sup_{s\in\mathbb{R}} \left| \frac{e^{is} - 1}{is} \right| \quad (5.11)$$

is square integrable. The result follows by Lebesgue dominated convergence.

Thus $M((F, f_1), \cdots, (F, f_n))\Psi(F) \in D_{\rho(f)}$ for all $f \in \mathfrak{J}$. Taking the vector space of finite linear combinations of such elements, we arrive at a common dense invariant domain for all of the $\rho(f)$.

Since $\Omega(F) \equiv 1$ is essentially bounded, Ω is in this domain. QED

Assume now that we have fixed on a common dense invariant domain D for all of the $\rho(f)$. Another assumption that can be made on the representation is the following.

Assumption 2: For all $\Psi, \Phi \in D$, $(\Psi, \rho(f)\Phi)$ is continuous in f.[35]

Lemma 4: If $f_i \to f$ and $g_j \to g$ in $\mathfrak{J}(\mathbb{R}^s)$, then $f_i \otimes g_j \to f \otimes g$ in $\mathfrak{J}(\mathbb{R}^{2s})$, where $(f \otimes g)(\mathbf{x}, \mathbf{y}) = f(\mathbf{x})g(\mathbf{y})$; i.e., "tensoring" is jointly continuous from $\mathfrak{J}(\mathbb{R}^s) \times \mathfrak{J}(\mathbb{R}^s)$ into $\mathfrak{J}(\mathbb{R}^{2s})$.

Proof: The proof is elementary and may be found in Ref. 5.

Theorem 12: Under Assumption 2, $\rho(f)\Psi$ is continuous in f for all $\Psi \in D$; i.e., the weak continuity of $\rho(f)$ with respect to domain D implies its strong continuity with respect to D.

Proof: Take $\Psi \in D$. Since D is invariant for the ρ's, Assumption 2 implies that $(\Psi, \rho(f_1)\rho(f_2)\Psi)$ is a tempered distribution in each variable separately. By Schwartz's nuclear theorem,[36] there exists a unique tempered distribution on $\mathfrak{J}(\mathbb{R}^{2s})$ which equals $(\Psi, \rho(f_1)\rho(f_2)\Psi)$ when evaluated at $f_1 \otimes f_2 \in \mathfrak{J}(\mathbb{R}^{2s})$.

Suppose then that $f_j \to f$ in $\mathfrak{J}(\mathbb{R}^s)$. Then

$$\|\rho(f_j)\Psi - \rho(f)\Psi\|^2 = (\Psi, \rho(f_j - f)\rho(f_j - f)\Psi) \to 0$$

since $(f_j - f) \otimes (f_j - f) \to 0$ in $\mathfrak{J}(\mathbb{R}^{2s})$ by Lemma 4. QED

Assumption 2 should be compared with the "continuity condition" for cylindrical measures, Eq. (A3). That condition states that if $\alpha(S_1, \cdots, S_m)$ is a bounded continuous function of m variables, then $\int_{\mathfrak{J}'} \alpha((F, f_1), \cdots, (F, f_m)) \, d\mu(F)$ is continuous in the variable $f_1, \cdots, f_m$.

With $\Omega \in D$, Assumption 2 extends the class of α's to include polynomials $P(S_1, \cdots, S_m)$; for

$$\int_{\mathfrak{J}'} P((F, f_1), \cdots, (F, f_m)) \, d\mu(F) = (\Omega, P(\rho(f_1), \cdots, \rho(f_m))\Omega),$$

which (by Lemma 4 and the nuclear theorem) is jointly continuous in $f_1, \cdots, f_m$.

Thus we may restate Assumption 2:

If $\alpha(\mathfrak{S}_1, \cdots, \mathfrak{S}_m)$ is a polynomial times a bounded continuous function of m real variables, then the integral $\int_{\mathfrak{J}'} \alpha((F,f_1), \cdots, (F,f_m))\, d\mu(F)$ is continuous in the variables $f_1, \cdots, f_m$.

Assumptions 3 and 4 apply to the representation V as well as to U. Assumption 3 simply states that $J(\mathbf{g})$ may be applied to the cyclic vector. If we let $\chi^{\mathbf{g}}(F) = [J(\mathbf{g})\Omega](F)$, Assumption 4 states that $\chi^{\mathbf{g}}$ is infinitely differentiable in a suitable sense. This strong assumption lets us apply polynomials in J to the desired domain of vectors.

Assumption 3: Let μ^{Ψ} be the measure defined by $d\mu^{\Psi}(F) = d\mu(\Psi^* F)$. Then, for all $\mathbf{g}$,

$$\lim_{t\to 0} \frac{1}{it}\left[\left(\frac{d\mu^{\varphi_t^{\mathbf{g}}}}{d\mu}(F)\right)^{\frac{1}{2}} - 1\right] \tag{5.12}$$

exists in $L^2_{\mu}(\mathfrak{J}')$.

Assumption 4: Let $\mathcal{O}_M(n)$ be the space of complex-valued C_∞ functions (of n real variables) which, together with all derivatives, are polynomially bounded. (The degree of the bounding polynomial can depend on the order of the derivative.) Then, for all $\mathbf{g}$, $\chi^{\mathbf{g}}(F) = \beta((F,f_1), \cdots, (F,f_n))$ a.e., for some n, $\beta \in \mathcal{O}_M(n)$, and $f_1, \cdots, f_n \in \mathfrak{J}$. Note that the number of variables n, as well as the function β and the set $f_1, \cdots, f_n$, may depend on $\mathbf{g}$.

Theorem 13: Let $U(f)V(\Psi)$ be a strongly continuous unitary representation of $\mathfrak{J} \wedge \mathfrak{K}$ in $\mathfrak{H}$; let $\Omega \in \mathfrak{H}$ be cyclic for $U(\mathfrak{J})$, and let $(\Omega, U(f)\Omega) = \int_{\mathfrak{J}'} e^{i(F,f)}\, d\mu(F)$ define a cylindrical measure μ in $\mathfrak{J}'$, quasi-invariant under $\mathfrak{K}$.

Let μ satisfy Assumptions 1–4 above.

Then in the representation of $\mathfrak{J} \wedge \mathfrak{K}$ in $L^2_{\mu}(\mathfrak{J}')$ defined by

$$U(f)\Psi(F) = e^{i(F,f)}\Psi(F)$$

and

$$V(\Psi)\Psi(F) = \Psi(\Psi^* F)\left(\frac{d\mu^{\Psi}}{d\mu}\right)^{\frac{1}{2}}(F),$$

there exists a common, dense, invariant domain for all $\rho(f)$ and $J(\mathbf{g})$. The cyclic vector $\Omega(F) \equiv 1$ is in this domain.

Proof: The domain D will be the linear subspace generated by

$$\{\beta((F,f_1), \cdots, (F,f_n)); \quad (\forall n)[\forall \beta \in \mathcal{O}_M(n)](\forall f_1, \cdots, f_n \in \mathfrak{J})\}.$$

Since $\Omega(F) \equiv 1$, $\Omega \in D$. Since $[U(f)\Omega](F) = e^{i(F,f)}$ and $e^{iS} \in \mathcal{O}_M(1)$, we have $U(f)\Omega \in D$; therefore, D is dense.

For all $f \in \mathfrak{J}$, $\rho(f)$ may be applied to elements of D, and $\rho(f)\Psi(F) = (F,f)\Psi(F)$ which is again in D (cf. Theorem 11). It remains to show that $J(\mathbf{g})$ can be applied to elements of D and that $J(\mathbf{g})D \subseteq D$.

Let $\beta((F,f_1), \cdots, (F,f_n)) \in D$. Where no confusion can result, we shall write simply $\beta(\cdots)$; the three dots take the place of $(F,f_1), \cdots, (F,f_n)$.

First, we find a certain sequence $\beta_i \in \mathcal{O}_M(n)$, such that β_i is bounded and satisfies a Lipschitz condition with $\beta_i(\cdots) \to \beta(\cdots)$ in $L^2_{\mu}(\mathfrak{J}')$. Since $\beta(\mathfrak{S}_1, \cdots, \mathfrak{S}_n)$, abbreviated $\beta(\mathfrak{S})$, is polynomially bounded, let us write $\beta(\mathfrak{S}) = P(\mathfrak{S})\beta_1(\mathfrak{S})$, where P is a polynomial in n variables and $\lim \beta_1(\mathfrak{S}) = 0$ as $|\mathfrak{S}| \to \infty$. Let $P_R(\mathfrak{S}) \in \mathcal{O}_M(n)$ be defined by $P_R(\mathfrak{S}) = \alpha_R(|\mathfrak{S}|)P(\mathfrak{S})$, where $\alpha_R(|\mathfrak{S}|) = 1$ for $|\mathfrak{S}| \leq R$, $\alpha_R(|\mathfrak{S}|) = 0$ for $|\mathfrak{S}| \geq R + \Delta R$, and $\alpha_R(|\mathfrak{S}|)$ is C_∞ in $\mathfrak{S}$, decreasing in $|\mathfrak{S}|$ for $R < |\mathfrak{S}| < R + \Delta R$. We can demand that

$$\sup_{\mathfrak{S}\in\mathbb{R}^n} |(\nabla\alpha_R)(\mathfrak{S})|$$

be bounded as R varies, since ΔR is not required to be small.

Then $|P(\mathfrak{S}) - P_R(\mathfrak{S})| \leq |P(\mathfrak{S})|$, and

$$|\nabla P(\mathfrak{S}) - \nabla P_R(\mathfrak{S})| \leq |\nabla P(\mathfrak{S})| + K|P(\mathfrak{S})|,$$

where

$$K = \max_R \sup_{\mathfrak{S}\in\mathbb{R}^n} |\nabla\alpha_R(\mathfrak{S})|.$$

Let $\beta_i(\mathfrak{S}) = P_{R_i}(\mathfrak{S})\beta_1(\mathfrak{S})$, where $R_i \to \infty$. It follows that β_i is bounded and satisfies the Lipschitz condition

$$|\beta_i(\mathfrak{S}_1) - \beta_i(\mathfrak{S}_2)| \leq K' \cdot |\mathfrak{S}_1 - \mathfrak{S}_2|, \tag{5.13}$$

where K, may be chosen independent of i, to be

$$\max_i \sup_{\mathfrak{S}\in\mathbb{R}^n} |\nabla\beta_i(\mathfrak{S})|.$$

Also, $\beta_i((F,f_1), \cdots, (F,f_n)) \to \beta(\cdots)$ in $L^2_{\mu}(\mathfrak{J}')$, since

$$\int_{\mathfrak{J}'} |\beta_i(\cdots) - \beta(\cdots)|^2\, d\mu(F)$$

$$\leq \sup_{\mathfrak{S}\in\mathbb{R}^n} |\beta_1(\mathfrak{S})|^2 \int_{\mathfrak{J}'} |P_{R_i}(\cdots) - P(\cdots)|^2\, d\mu(F)$$

$$\leq \text{const} \times \int_{\{F \mid \sum_{j=1}^n |(F,f_j)|^2 > R_i^2\}} |P(\cdots)|^2\, d\mu(F). \tag{5.14}$$

As $P(\cdots)$ is square integrable by Assumption 1, (5.14) approaches zero as $R_i \to \infty$.

Our goal is to show $\beta(\cdots) \in D_{J(\mathbf{g})}$, and we have constructed $\beta_i(\cdots) \to \beta(\cdots)$ in $L^2_\mu(\mathfrak{J}')$. Next we show that $\beta_i(\cdots) \in D_{J(\mathbf{g})}$.

In fact,

$$\lim_{t\to 0} \frac{V(\varphi^{\mathbf{g}}_t) - I}{it}\beta_i(\cdots)$$

$$= \lim_{t\to 0}\left(\frac{\beta_i((F, f_1\circ\varphi^{\mathbf{g}}_t), \cdots, (F, f_n\circ\varphi^{\mathbf{g}}_t)) - \beta_i(\cdots)}{it}\right.$$

$$+ \beta_i(\cdots)\left\{\left[\left(\frac{d\mu^{\varphi_t^{\mathbf{g}}}}{d\mu}(F)\right)^{\frac{1}{2}} - 1\right]\Big/ it\right\}$$

$$+ [\beta_i((F, f_1\circ\varphi^{\mathbf{g}}_t), \cdots, (F, f_n\circ\varphi^{\mathbf{g}}_t)) - \beta_i(\cdots)]$$

$$\left.\times\left\{\left[\left(\frac{d\mu^{\varphi_t^{\mathbf{g}}}}{d\mu}(F)\right)^{\frac{1}{2}} - 1\right]\Big/ it\right\}\right). \tag{5.15}$$

We must show that the limit of each term in (5.15) exists in $L^2_\mu(\mathfrak{J}')$.

Now (5.13) states that

$$\left|\frac{\beta_i((F, f_1\circ\varphi^{\mathbf{g}}_t), \cdots, (F, f_n\circ\varphi^{\mathbf{g}}_t)) - \beta_i(\cdots)}{it}\right|^2$$

$$\leq C^2\sum_{j=1}^{n}\left|\left(F, \frac{f_j\circ\varphi^{\mathbf{g}}_t - f_j}{t}\right)\right|^2. \tag{5.16}$$

But $(F, (f_j\circ\varphi^{\mathbf{g}}_t - f_j)/t)$ converges pointwise to $(F, \mathbf{g}\cdot\nabla f_j)$ by the continuity of F and in $L^2_\mu(\mathfrak{J}')$ to $(F, \mathbf{g}\cdot\nabla f_j)$ by Assumption 2.[37] The Vitali convergence theorem, which is a slight extension of Lebesgue dominated convergence, now suffices to conclude that the first term of (5.15) converges to its pointwise limit

$$\frac{1}{i}\sum_{j=1}^{n}\left(\frac{\partial\beta_i}{\partial\mathfrak{S}_j}(\cdots)\right)(F, \mathbf{g}\cdot\nabla f_j). \tag{5.17}$$

The Vitali convergence theorem may be found in Appendix B.

The second term in (5.15) converges because $\beta_i(\cdots)$ is bounded, while

$$\lim_{t\to 0}\left[\left(\frac{d\mu^{\varphi_t^{\mathbf{g}}}}{d\mu}(F)\right)^{\frac{1}{2}} - 1\right]\Big/ it = \chi^{\mathbf{g}}(F) \tag{5.18}$$

by Assumption 3.

The third term in (5.15) is bounded in the square of its norm by

$$C^2\sum_{j=1}^{n}\left\|\frac{1}{i}\rho\left(\frac{f_j\circ\varphi^{\mathbf{g}}_t - f_j}{t}\right)[V(\varphi^{\mathbf{g}}_t) - 1]\Omega\right\|^2,$$

which converges to zero by arguing as in the proof of Theorem 10(d).

Thus $\beta_i(\cdots) \in D_{J(\mathbf{g})}$ for all $\mathbf{g}$, and

$$J(\mathbf{g})\beta_i(\cdots) = \frac{1}{i}\sum_{j=1}^{n}\frac{\partial\beta_i}{\partial\mathfrak{S}_j}(\cdots)(F, \mathbf{g}\cdot\nabla f_j)$$

$$+ \beta_i(\cdots)\chi^{\mathbf{g}}(F). \tag{5.19}$$

Now we show that as $i \to \infty$, (5.19) converges in $L^2_\mu(\mathfrak{J}')$ to

$$\frac{1}{i}\sum_{j=1}^{n}\frac{\partial\beta}{\partial\mathfrak{S}_j}(\cdots)(F, \mathbf{g}\cdot\nabla f_j) + \beta(\cdots)\chi^{\mathbf{g}}(F). \tag{5.20}$$

In fact,

$$\left|\frac{\partial\beta}{\partial\mathfrak{S}_j}(\cdots) - \frac{\partial\beta_i}{\partial\mathfrak{S}_j}(\cdots)\right|$$

$$\leq \left|\frac{\partial(P - P_{R_i})}{\partial\mathfrak{S}_j}\beta_1\right| + \left|(P - P_{R_i})\frac{\partial\beta_1}{\partial\mathfrak{S}_j}\right|$$

$$\leq \text{const} \times |\nabla P(\cdots)| + \text{const} \times |P(\cdots)|. \tag{5.21}$$

Thus

$$\frac{1}{i}\left(\frac{\partial\beta}{\partial\mathfrak{S}_j}(\cdots) - \frac{\partial\beta_i}{\partial\mathfrak{S}_j}(\cdots)\right)(F, \mathbf{g}\cdot\nabla f_j) \tag{5.22}$$

is bounded in absolute value by an element of $L^2_\mu(\mathfrak{J}')$, independently of i. Since, for each point F, the limit as $i \to \infty$ of (5.22) is zero, we have Lebesgue dominated convergence, and the limit is zero in $L^2_\mu(\mathfrak{J}')$.

Finally, $\beta_i(\cdots)\chi^{\mathbf{g}}(F) \to \beta(\cdots)\chi^{\mathbf{g}}(F)$ via Lebesgue dominated covergence and the fact that

$$|\beta_i(\cdots)\chi^{\mathbf{g}}(F)| \leq |\beta(\cdots)\chi^{\mathbf{g}}(F)|.$$

Since $J(\mathbf{g})$ is a self-adjoint operator, its graph is closed. Hence $\beta(\cdots) \in D_{J(\mathbf{g})}$, and $J(\mathbf{g})\beta(\cdots)$ is given by (5.20). It is clear from (5.20) that $J(\mathbf{g})D \subseteq D$.

QED

In Theorem 10(e), the relations among the $J(\mathbf{g})$ were only recovered up to commutation with ρ. Given the domain D of Theorem 13, it remains to recover them on the vector Ω.

Theorem 14: Under the assumptions of Theorem 13,

(a) $J(\lambda\mathbf{g}) = \lambda J(\mathbf{g})$,

(b) $J(\mathbf{g}_1 + \mathbf{g}_2) = J(\mathbf{g}_1) + J(\mathbf{g}_2)$,

(c) $[J(\mathbf{g}_1), J(\mathbf{g}_2)] = iJ(\mathbf{g}_2\cdot\nabla\mathbf{g}_1 - \mathbf{g}_1\cdot\nabla\mathbf{g}_2)$ on the whole domain D.

Proof: By Theorem 10(e), it is sufficient to prove (a)–(c) on Ω. Let $M_{\chi^{\mathbf{g}}}$ be the operation of multiplication by $\chi^{\mathbf{g}}(F)$, defined on D:

(a) Let $J'(\mathbf{g}) = J(\mathbf{g}) - M_{\chi^{\mathbf{g}}}$. Then, for $\Psi \in D$, $J'(\lambda\mathbf{g})\Psi = \lambda J'(\mathbf{g})\Psi$, whence

$$(\Psi, J'(\lambda\mathbf{g})^*\Omega) = (\Psi, \lambda J'(\mathbf{g})^*\Omega).$$

As $\chi^{\mathbf{g}}(F)$ is purely imaginary (a.e.),

$$J(\mathbf{g}) = \tfrac{1}{2}[J'(\mathbf{g}) + J'(\mathbf{g})^*] \quad \text{and} \quad J(\lambda\mathbf{g})\Omega = \lambda J(\mathbf{g})\Omega.$$

(b) The argument is identical.

(c) Let $A = [J(\mathbf{g}_1), J(\mathbf{g}_2)] - iJ(\mathbf{g}_2\cdot\nabla\mathbf{g}_1 - \mathbf{g}_1\cdot\nabla\mathbf{g}_2)$. A is defined on D and commutes with ρ; therefore

$A = M_{(A\Omega)(F)}$ on D. Now $A^* = -A$; thus $(A\Omega)(F)$ is purely imaginary (a.e.). But $\chi^g(F)$ is purely imaginary, whence $(A\Omega)(F)$ is pure real (a.e.). Therefore $(A\Omega)(F) = 0$. QED

Theorem 15: $\rho(f)$ is essentially self-adjoint on the domain D of Theorem 13.

Proof: Since $U(tf)D \subseteq D$, the result follows as in the proof of Lemma 3, Sec. 3. QED

B. Remarks

(1) It is well known in the folklore of axiomatic field theory that when the fields can be applied to the vacuum state, they can generally be recovered on a dense domain. The above development restates the analogous result for nonrelativistic currents, in the Gel'fand–Vilenkin formalism.

It would be interesting to investigate ways in which the Assumptions 1–4 might be weakened.

(2) One may contrast the semidirect product case above, with the parallel development for the Weyl group.[5] A representation of the Weyl group satisfies, we recall, $V(g)U(f) = e^{i(f,g)}U(f)V(g)$, for $f, g \in \mathfrak{I}$, $U(f) = e^{i\varphi(f)}$, $V(g) = e^{i\pi(g)}$. An element g acts on $\mathfrak{I}'$ by translation, i.e., $F \to F + F_g$, where $(F_g, f) \equiv (f, g)$, and one obtains, in $L^2_\mu(\mathfrak{I}')$,

$$U(f)\Psi(F) = e^{i(F,f)}\Psi(F), \tag{5.23}$$

$$V(g)\Psi(F) = \Psi(F + F_g)\left(\frac{d\mu^g}{d\mu}(F)\right)^{\frac{1}{2}}, \tag{5.24}$$

where $d\mu^g(F) = d\mu(F + F_g)$ and the "multiplier" has been set equal to one.[29]

Then the assumption that $\varphi(f)\Psi$ is continuous in f for $\Psi \in D$ is no longer necessary for the recovery of the commutation relations.

Also, the *continuity* of the linear functional F, which was used in the proof of Theorem 13, is not needed for the analogous result in the case of the Weyl group. In fact, the Gel'fand spectral theory discussed in Sec. 4, in which the measure space contains linear functionals which are *not* necessarily continuous, is an adequate framework for constructing a common dense invariant domain for the $\varphi(f)$ and the $\pi(g)$ in analogy with Theorem 13. Thus, one advantage of the Gel'fand–Vilenkin formalism is that it enables us to extend this treatment to a wider class of commutation relations.

It is not difficult to verify that Gaussian measures in $\mathfrak{I}'$, which lead to "free field" representations of the Weyl group,[29] satisfy the assumptions on μ in Theorem 13.

(3) Let us show explicitly that Assumptions 1–4 on μ in Theorem 13 are satisfied for the 1-particle Fock representation. First of all,

$$\int_{\mathfrak{I}'} |(F, f_1)|^2 \cdots |(F, f_n)|^2 \, d\mu(F)$$

$$= \pi^{-\frac{1}{2}s} \int_{\mathbb{R}^s} |f_1(\mathbf{x})|^2 \cdots |f_n(\mathbf{x})|^2 \, e^{-|\mathbf{x}|^2} \, d\mathbf{x} < \infty. \tag{5.25}$$

Secondly, if α is a polynomial times a bounded continuous function of n variables, then

$$\int_{\mathfrak{I}'} \alpha((F, f_1), \cdots, (F, f_n)) \, d\mu(F)$$

$$= \pi^{-\frac{1}{2}s} \int_{\mathbb{R}^s} \alpha(f_1(\mathbf{x}), \cdots, f_n(\mathbf{x})) e^{-|\mathbf{x}|^2} \, d\mathbf{x}. \tag{5.26}$$

But, if

$$\lim_{i \to \infty} f_k^i = f_k \quad \text{in } \mathfrak{I}, \quad k = 1, \cdots, n,$$

then $\alpha(f_1^i(\mathbf{x}), \cdots, f_n^i(\mathbf{x})) \to \alpha(f_1(\mathbf{x}), \cdots, f_n(\mathbf{x}))$ uniformly in $\mathbf{x}$, and (5.26) is continuous in $(f_1, \cdots, f_n)$.

Thirdly, we have

$$\chi^g(F_\mathbf{x}) = [J(\mathbf{g})\Omega](F_\mathbf{x})$$

$$= \frac{1}{i} \frac{\partial}{\partial t} \frac{\exp(-\frac{1}{2}|\boldsymbol{\varphi}_t^g(\mathbf{x})|^2)}{\exp(-\frac{1}{2}|\mathbf{x}|^2)} \left[\det\left(\frac{\partial(\boldsymbol{\varphi}_t^g)^l}{\partial x^k}\right)\right]^{\frac{1}{2}}\Bigg|_{t=0}, \tag{5.27}$$

which equals

$$i^{-1}[-\mathbf{x} \cdot \mathbf{g}(\mathbf{x}) + \tfrac{1}{2}\nabla \cdot \mathbf{g}(\mathbf{x})], \tag{5.28}$$

using (3.10). It is not difficult to check that (5.27) converges to (5.28) uniformly in $\mathbf{x}$, which *a fortiori* implies convergence in $L^2_\mu(\mathfrak{I}')$.

Finally, with

$$\beta(\mathcal{S}_1, \cdots, \mathcal{S}_s, \mathcal{S}_{s+1}, \cdots, \mathcal{S}_{2s}) = i^{-1}\left(-\sum_{j=1}^{s} \mathcal{S}_j + \tfrac{1}{2}\sum_{j=s+1}^{2s} \mathcal{S}_j\right),$$

we have almost everywhere

$$\chi^g(F) = \beta\Big((F, x^1 g^1), \cdots, (F, x^s g^s), \left(F, \frac{\partial g^1}{\partial x^1}\right), \cdots, \left(F, \frac{\partial g^s}{\partial x^s}\right)\Big), \tag{5.29}$$

fulfilling Assumption 4.

It is not difficult to convince oneself that the n-particle Fock representation, for which the measure is concentrated on

$$\{F \in \mathfrak{I}' \mid F = F_{\mathbf{x}_1} + F_{\mathbf{x}_2} + \cdots + F_{\mathbf{x}_n}; \mathbf{x}_j \neq \mathbf{x}_k\},$$

also satisfies Assumptions 1–4.

C. Application to the Theory in a Box

In this subsection, we complete the discussion of the theory in a box, by recovering Eqs. (4.55)–(4.57) on the domain D' defined in Sec. 4.

Theorem 16: Let μ be a cylindrical measure on the dual of $C_\infty(T^s)$, quasi-invariant under the action of $\mathfrak{W}$ and satisfying Assumptions 1–4.

Then, in the representation of $C_\infty(T^s) \wedge \mathfrak{W}$ which μ defines, $\rho_{\mathbf{m}}$ and $J_{\mathbf{n}}$ can be recovered on a common dense invariant domain D. The domain D includes the set

$$D' = \{\Psi'(\Lambda) = \beta(\mathrm{Re}\,\lambda_{\mathbf{m}_1}, \mathrm{Im}\,\lambda_{\mathbf{m}_1}, \cdots, \mathrm{Re}\,\lambda_{\mathbf{m}_n}, \mathrm{Im}\,\lambda_{\mathbf{m}_n}); (\forall n)[\forall \beta \in \mathcal{O}_M(2n)](\forall \mathbf{m}_1, \cdots, \mathbf{m}_n)\}.$$

D' is a domain of essential self-adjointness for $\rho^+_{\mathbf{m}}$ and $\rho^-_{\mathbf{m}}$.

The $J^k_{\mathbf{m}}$ can be represented on D' in $L^2_\mu(\{\Lambda\})$ by

$$J^k_{\mathbf{m}} = -\frac{\pi}{a}\frac{1}{[(2a)^s]^{\frac{1}{2}}}\sum_{\ell=-\infty}^{\infty}\ell^k\lambda_{\mathbf{m}+\ell}\frac{\partial}{\partial\lambda_\ell} + M_{(J_{\mathbf{m}}{}^k\Omega)(\Lambda)}. \tag{5.30}$$

In particular, the sum in (5.30) is a well-defined operator on D'.

Proof: The common, dense, invariant domain D is constructed exactly as in Theorem 13. Thus D is the linear subspace of $L^2_\mu(C_\infty(T^s)')$ generated by

$$\{\beta((F,f_1),\cdots,(F,f_n)); (\forall n)[\forall\beta\in\mathcal{O}_M(n)][\forall f_1,\cdots,f_n\in C_\infty(T^s)]\}.$$

Choosing the $f_1,\cdots,f_n$ from among the cosine and sine functions

$$[(2a)^s]^{-\frac{1}{2}}\cos[-\mathbf{m}\cdot\mathbf{x}(\pi/a)]$$

and

$$[(2a)^s]^{-\frac{1}{2}}\sin[-\mathbf{m}\cdot\mathbf{x}(\pi/a)]$$

and identifying F with the infinite sequence $\Lambda = \{\lambda_{\mathbf{m}}\}$, where

$$\lambda_{\mathbf{m}} = [(2a)^s]^{-\frac{1}{2}}\int F(\mathbf{x})e^{-i\mathbf{m}\cdot\mathbf{x}\pi/a}\,d\mathbf{x}, \tag{5.31}$$

one obtains the smaller domain D'.

The 1-parameter unitary groups (4.53) and (4.54), which map D into itself, likewise map D' into itself. Thus (as in the proof of Lemma 3), D' is a domain of essential self-adjointness for $\rho^+_{\mathbf{m}}$ and $\rho^-_{\mathbf{m}}$.

Now, according to Theorem 13, the operator $J(\mathbf{g})$ is represented on D by

$$J(\mathbf{g})\beta((F,f_1),\cdots,(F,f_n)) = \frac{1}{i}\sum_{j=1}^{n}\frac{\partial\beta}{\partial\mathcal{S}_j}(\cdots)(F,\mathbf{g}\cdot\nabla f_j) + \beta(\cdots)[J(\mathbf{g})\Omega](F). \tag{5.32}$$

Thus on D', with $f_1,\cdots,f_{2n}$ and g^k all of the form of the above cosine and sine functions, we obtain

$$\begin{aligned}J^k_{\mathbf{m}}&\beta(\mathrm{Re}\,\lambda_{\mathbf{m}_1},\mathrm{Im}\,\lambda_{\mathbf{m}_1},\cdots,\mathrm{Re}\,\lambda_{\mathbf{m}_n},\mathrm{Im}\,\lambda_{\mathbf{m}_n})\\ &= \frac{1}{i}\sum_{j=1}^{n}\frac{\partial\beta}{\partial\mathcal{S}_{2j-1}}(\mathrm{Re}\,\lambda_{\mathbf{m}_1},\cdots,\mathrm{Im}\,\lambda_{\mathbf{m}_n})\\ &\quad\times\left(F,f_{(\mathbf{m})}\cdot\left(m_j^k\frac{\pi}{a}\right)(2a)^{-\frac{1}{2}s}\sin\left(-\mathbf{m}_j\cdot\mathbf{x}\frac{\pi}{a}\right)\right)\\ &\quad+\frac{1}{i}\sum_{j=1}^{n}\frac{\partial\beta}{\partial\mathcal{S}_{2j}}(\mathrm{Re}\,\lambda_{\mathbf{m}_1},\cdots,\mathrm{Im}\,\lambda_{\mathbf{m}_n})\\ &\quad\times\left(F,f_{(\mathbf{m})}\cdot\left(-m_j^k\frac{\pi}{a}\right)(2a)^{-\frac{1}{2}s}\cos\left(-\mathbf{m}_j\cdot\mathbf{x}\frac{\pi}{a}\right)\right)\\ &\quad+\beta(\mathrm{Re}\,\lambda_{\mathbf{m}_1},\cdots,\mathrm{Im}\,\lambda_{\mathbf{m}_n})(J^k_{\mathbf{m}}\Omega)(\Lambda).\end{aligned} \tag{5.33}$$

But

$$(F,f_{(\mathbf{m})}\cdot(m_j^k\pi/a)(2a)^{-\frac{1}{2}s}\sin(-\mathbf{m}_j\cdot\mathbf{x}\pi/a)) = (2a)^{-\frac{1}{2}s}m_j^k\pi/a\cdot(2i)^{-1}(\lambda_{\mathbf{m}+\mathbf{m}_j}-\lambda_{\mathbf{m}-\mathbf{m}_j}) \tag{5.34}$$

and

$$(F,f_{(\mathbf{m})}\cdot(-m_j^k\pi/a)(2a)^{-\frac{1}{2}s}\cos(-\mathbf{m}_j\cdot\mathbf{x}\pi/a)) = -(2a)^{-\frac{1}{2}s}m_j^k\pi/a\cdot\tfrac{1}{2}(\lambda_{\mathbf{m}+\mathbf{m}_j}+\lambda_{\mathbf{m}-\mathbf{m}_j}). \tag{5.35}$$

With

$$\frac{\partial}{\partial\lambda_{\mathbf{m}_j}} = \frac{1}{2}\left(\frac{\partial}{\partial(\mathrm{Re}\,\lambda_{\mathbf{m}_j})}+\frac{1}{i}\frac{\partial}{\partial(\mathrm{Im}\,\lambda_{\mathbf{m}_j})}\right), \tag{5.36}$$

we obtain

$$\begin{aligned}J^k_{\mathbf{m}}&\beta(\mathrm{Re}\,\lambda_{\mathbf{m}_1},\cdots,\mathrm{Im}\,\lambda_{\mathbf{m}_n})\\ &= -\frac{1}{[(2a)^s]^{\frac{1}{2}}}\frac{\pi}{a}\sum_{j=1}^{n}m_j^k(\lambda_{\mathbf{m}+\mathbf{m}_j})\frac{\partial\beta}{\partial\lambda_{\mathbf{m}_j}}\\ &\quad+\frac{1}{[(2a)^s]^{\frac{1}{2}}}\frac{\pi}{a}\sum_{j=1}^{n}m_j^k(\lambda_{\mathbf{m}-\mathbf{m}_j})\frac{\partial\beta}{\partial\bar\lambda_{\mathbf{m}_j}}+\beta(J^k_{\mathbf{m}}\Omega)(\Lambda)\\ &= -\frac{1}{[(2a)^s]^{\frac{1}{2}}}\frac{\pi}{a}\sum_{\ell=-\infty}^{\infty}\ell^k\lambda_{\mathbf{m}+\ell}\frac{\partial\beta}{\partial\lambda_\ell}+\beta(J^k_{\mathbf{m}}\Omega)(\Lambda),\end{aligned} \tag{5.37}$$

which is the desired result. QED

6. CONCLUSION

We have examined a theory in which nonrelativistic quantum mechanics is rewritten in terms of the algebra of currents ρ and J, and have found that we are able to exponentiate the current algebra successfully. Instead of representations of the current algebra by (unbounded) self-adjoint operators, we can then consider unitary representations of the corresponding group.

The Gel'fand–Vilenkin formalism provides a reasonably unified framework in which to describe such representations. The n-particle representations of

ordinary quantum mechanics have been recovered—the distinction between n particles satisfying Bose statistics and n particles satisfying Fermi statistics then appears in the "multiplier" χ, rather than in the cylindrical measure itself.

Conditions on the measure μ have been obtained which suffice to recover all of the infinitesimal generators on a common, dense domain in the Hilbert space. We have described the theory of N particles in a box in this formalism and have succeeded in giving mathematically rigorous meaning to the formal representations studied by Grodnik and Sharp.

The present reformulation of nonrelativistic quantum mechanics will hopefully provide guidelines for the study of relativistic models written entirely in terms of local currents,[38] as well as proving suitable in its own right for the study of nonrelativistic problems involving infinitely many particles.[39]

ACKNOWLEDGMENTS

The author wishes to thank A. S. Wightman for his constant advice and encouragement in carrying out this research and for his many helpful suggestions. I am indebted for valuable discussions to A. Dicke, J. Grodnik, R. Powers, D. H. Sharp, B. Simon, and G. Svetlichny. The research was carried out under a National Science Foundation Fellowship.

APPENDIX A

This section is devoted to the presentation of basic definitions and results, in the Gel'fand–Vilenkin formalism.[29]

Definition 1: Let $\mathcal{H}_1$ and $\mathcal{H}_2$ be Hilbert spaces. The linear operator $A:\mathcal{H}_1 \to \mathcal{H}_2$ is called *completely continuous* if and only if the image of any bounded set is relatively compact. Equivalently, A maps weakly convergent sequences into strongly convergent sequences.

Proposition 1: Let $A:\mathcal{H}_1 \to \mathcal{H}_2$ be completely continuous. Then A can be written $A = UT$, where T is a positive-definite completely continuous operator in $\mathcal{H}_1$ and where U is an isometry from the range of T into $\mathcal{H}_2$.

A self-adjoint completely continuous operator T has the property that one can choose an orthonormal basis $\Psi_1, \Psi_2, \cdots$ of eigenvectors of T, $T\Psi_n = \lambda_n \Psi_n$, with $\lim \lambda_n = 0$ as $n \to \infty$.

Definition 2: The completely continuous operator $A = UT$ is of *Hilbert–Schmidt type* if and only if $\sum_{n=1}^{\infty} \lambda_n^2 < \infty$, where the λ_n are the eigenvalues of T. For A to be of Hilbert–Schmidt type, it is necessary and sufficient that there exist an orthonormal basis Ψ_n for which $\sum_{n=1}^{\infty} \|A\Psi_n\|^2$ converges.

Definition 3: The completely continuous operator $A = UT$ is called *nuclear*, or said to be *of trace class*, if and only if $\sum_{n=1}^{\infty} \lambda_n < \infty$, where the λ_n are the eigenvalues of T. Evidently every nuclear operator is of Hilbert–Schmidt type.

Definition 4: Let $\mathcal{M}$ be a countably normed space. $\mathcal{M}$ is called *countably Hilbert* if and only if each norm has the form $\|f\|_n = [(f,f)_n]^{\frac{1}{2}}$, where the $(\ ,\ )_n$ are a compatible family of scalar products on $\mathcal{M}$. A countably normed or countably Hilbert space is understood to be complete. Let $\mathcal{M}_n$ denote the completion of $\mathcal{M}$ with respect to the scalar product $(\ ,\ )_n$. $\mathcal{M}$ may be written $\bigcap_{n=1}^{\infty} \mathcal{M}_n$.

Proposition 2: Let $\mathcal{M}$ be a countably Hilbert space. Then its continuous dual $\mathcal{M}'$ may be written $\bigcup_{n=1}^{\infty} \mathcal{M}_n'$, where $\mathcal{M}_n'$ is the dual of the Hilbert space $\mathcal{M}_n$.

In $\mathcal{M}_n'$ we define the norm

$$\|F\|_{-n} = \sup_{\|f\|=1} (F,f). \tag{A1}$$

Since $\mathcal{M}_n'$ is a Hilbert space, $\|F\|_{-n}$ is defined by a scalar product in $\mathcal{M}_n'$.

Proposition 3: A countably Hilbert space is *reflexive*; i.e., $\mathcal{M}'' = \mathcal{M}$.

Definition 5: Let $\mathcal{M}$ be a countably Hilbert space. If $m \leq n$, then, $\forall f \in \mathcal{M}$, $(f,f)_m \leq (f,f)_n$, since the scalar products define a compatible system of norms. Thus we can define $T_m^n: \mathcal{M}_n \to \mathcal{M}_m$ as follows—let $T_m^n(f) = f$ for $f \in \mathcal{M}$, and extend T_m^n to all of $\mathcal{M}_n$ by continuity. The space $\mathcal{M}$ is called *nuclear* if and only if for all n, there exists an $n \geq m$ such that T_m^n is nuclear.

Proposition 4: The Schwartz space $\mathfrak{I}$ is a nuclear space.

We continue with more definitions and results from Gel'fand and Vilenkin.

Definition 6: Let $\mathcal{M}$ be a locally convex topological vector space, and $\mathcal{N}$ a finite-dimensional subspace of $\mathcal{M}$. Let $\mathcal{N}^0 \subseteq \mathcal{M}'$ be defined by

$$\mathcal{N}^0 = \{F \in \mathcal{M}' \mid F \mid \mathcal{N} = 0\}. \tag{A2}$$

$\mathcal{N}^0$ is called the *annihilator* of $\mathcal{N}$. The quotient space $\mathcal{M}'/\mathcal{N}^0$ may be identified with $\mathcal{N}'$, the adjoint space of $\mathcal{N}$.

Definition 7: Let $A \subseteq \mathcal{M}'/\mathcal{N}^0$. Then

$$X_A^{\mathcal{N}} = \{F \in \mathcal{M}' \mid F + \mathcal{N}^0 \in A\}$$

is called the *cylinder set* with *base* A and *generating subspace* $\mathcal{N}^0$.

Definition 8: Let

$$n = \dim \mathcal{N} = \dim \mathcal{N}' = \dim \mathcal{M}'/\mathcal{N}^0.$$

The cylinder set X_A has *Borel base* if A is Borel when regarded as a subset of $\mathbb{R}^n$. The family of cylinder sets with Borel base forms an *algebra* of sets.

Definition 9: The *measurable sets* in $\mathcal{M}'$ are the elements of the *σ-algebra* generated by the cylinder sets with Borel base.

Definition 10: A *cylindrical measure* in $\mathcal{M}'$ is a real-valued function μ defined on the algebra of cylinder sets with Borel base and satisfying (a) $0 \leq \mu(X) \leq 1$, $\forall X$, (b) $\mu(\mathcal{M}') = 1$, and (c) if X is the countable union of mutually disjoint cylinder sets X_i with Borel base and having a common generating subspace $\mathcal{N}^0$, then $\mu(X) = \sum_{i=1}^{\infty} \mu(X_i)$.

Definition 11: A cylindrical measure μ satisfies the *continuity condition* if and only if, for any bounded continuous function $\alpha(\delta_1, \cdots, \delta_m)$ of m real variables, the function

$$I(f_1, \cdots, f_m) = \int_{\mathcal{M}'} \alpha((F, f_1), \cdots, (F, f_m))\, d\mu(F) \tag{A3}$$

is sequentially continuous in $f_1, \cdots, f_m \in \mathcal{M}$. (In countably normed spaces, sequential and ordinary continuity are equivalent.)

Definition 12: A cylindrical measure μ is *countably additive* if and only if for any cylinder set X which is the union of countably many mutually disjoint cylinder sets X_i, $\mu(X) = \sum_{i=1}^{\infty} \mu(X_i)$.

Proposition 5: A countably additive cylindrical measure can be extended to a countably additive measure on the σ-algebra generated by the cylinder sets with Borel base. Such a measure will also be called a cylindrical measure.

Proposition 6: Let $\mathcal{M}$ be a nuclear space. Then any cylindrical measure μ in $\mathcal{M}'$, satisfying the continuity condition, is countably additive.

Definition 13: Let μ be a cylindrical measure in $\mathcal{M}'$. The *Fourier transform* of μ is the (nonlinear) functional

$$L(f) = \int_{\mathcal{M}'} e^{i(F,f)}\, d\mu(F). \tag{A4}$$

Definition 14: The functional $L(f)$ on $\mathcal{M}$ is called *positive definite* if and only if, $\forall f_1, \cdots, f_m \in \mathcal{M}$ and $\forall \lambda_1, \cdots, \lambda_m \in \mathbb{C}$,

$$\sum_{j,k=1}^{m} \bar{\lambda}_k \lambda_j L(f_j - f_k) \geq 0. \tag{A5}$$

Proposition 7: The functional $L(f)$ on $\mathcal{M}$ is the Fourier transform of a cylindrical measure on $\mathcal{M}'$ if and only if $L(f)$ is positive definite, (sequentially) continuous, and $L(0) = 1$.

Example: Let $B(f, g)$ be a scalar product in $\mathcal{M}$, continuous in the arguments f, g. Let $\mathcal{N}$ be an n-dimensional subspace of $\mathcal{M}$. Define a measure $\mu_{\mathcal{N}}$ in $\mathcal{N}$ by

$$\mu_{\mathcal{N}}(X) = \frac{1}{(2\pi)^{n/2}} \int_X e^{-\frac{1}{2}B(f,f)}\, df, \tag{A6}$$

where df is the Lebesgue measure in $\mathcal{N}$ corresponding to the scalar product $B(f, g)$. Since $\mathcal{N} \cong \mathcal{M}'/\mathcal{N}^0$, we see that $\mu_{\mathcal{N}}$ defines a measure $\nu_{\mathcal{N}}$ in $\mathcal{M}'/\mathcal{N}^0$. The $\nu_{\mathcal{N}}$ in turn defines a cylindrical measure μ in $\mathcal{M}'$, by the formula $\mu(X_A^{\mathcal{N}}) = \nu_{\mathcal{N}}(A)$. μ satisfies the continuity condition. If $\mathcal{M}$ is a nuclear space, then μ is countably additive. μ is called the *Gaussian measure* in $\mathcal{M}'$ defined by the scalar product $B(f, g)$.

The Fourier transform of μ is

$$L(f) = e^{-\frac{1}{2}B(f,f)}. \tag{A7}$$

Definition 15: Let $\mathcal{M}$ be a nuclear space in which there is defined a scalar product (f, g). Let $\mathcal{H}$ denote the completion of $\mathcal{M}$ with respect to the scalar product. Then $\mathcal{M} \subseteq \mathcal{H} \subseteq \mathcal{M}'$ is called a *rigged Hilbert space.*

Definition 16: Let $\mathcal{M} \subseteq \mathcal{H} \subseteq \mathcal{M}'$ be a rigged Hilbert space. For $g \in \mathcal{M}$, define $F_g \in \mathcal{M}'$ by $(F_g, f) = (g, f)$. Thus $\mathcal{M}$ is embedded in $\mathcal{M}'$; in fact, $\mathcal{M}$ is dense in $\mathcal{M}'$. Now the cylindrical measure μ in $\mathcal{M}'$ is *quasi-invariant* (for $\mathcal{M}$) if and only if, $\forall g \in \mathcal{M}$, $\mu(X) = 0$ implies $\mu(X + F_g) = 0$ for every measurable set X.

Proposition 8: Let $\mathcal{M} \subseteq \mathcal{H} \subseteq \mathcal{M}'$ be a rigged Hilbert space, with $B(f, g)$ the scalar product in $\mathcal{H}$. Then the Gaussian measure μ in $\mathcal{M}'$ defined by $B(f, g)$ is quasi-invariant, i.e., $\mu(X + F_g) = 0$ if and only if $\mu(X) = 0$, where $(F_g, f) = B(g, f)$.

APPENDIX B

The Lebesgue dominated convergence theorem used in Sec. 5 is a corollary of a more general result, the Vitali convergence theorem, which was also necessary to us in demonstrating Eq. (5.15).

Vitali Convergence Theorem: Let (X, Σ, μ) be a measure space, and let $\Psi_i(F) \in L^p(X, \Sigma, \mu)$ with $\lim \Psi_i(F) = \Psi(F)$ as $i \to \infty$. Then $\Psi(F)$ is an element of $L^p(X, \Sigma, \mu)$ and $\|\Psi_i - \Psi\|_p \to 0$ if and only if: (a) For all $E_j \in \Sigma$ such that $\lim \mu(E_j) = 0$ as $j \to \infty$,

$$\lim_{j \to \infty} \int_{E_j} |\Psi_i(F)|^p \, d\mu(F) = 0$$

uniformly in i; (b) for all $\epsilon > 0$, there exists $E_\epsilon \in \Sigma$ such that

$$\mu(E_\epsilon) < \infty \quad \text{and} \quad \int_{X-E} |\Psi_i(F)|^p \, d\mu(F) < \epsilon.$$

Condition (b) is trivially satisfied if $\mu(X) < \infty$.

Corollary 1 (Lebesgue Dominated Convergence): Suppose that $\Psi_i(F) \to \Psi(F)$ almost everywhere and $|\Psi_i(F)| \leq |\Phi(F)|$ almost everywhere, $\Phi \in L^p(X, \Sigma, \mu)$. Then $\Psi \in L^p(X, \Sigma, \mu)$ and $\|\Psi_i - \Psi\|_p \to 0$.

Corollary 2: Suppose that $\Psi_i(F) \to \Psi(F)$ almost everywhere and $|\Psi_i(F)| \leq |\Phi_i(F)|$ almost everywhere, with $\Phi_i(F) \to \Phi(F)$ almost everywhere, that Φ_i, $\Phi \in L^p(x, \Sigma, \mu)$, and that $\|\Phi_i - \Phi\|_p \to 0$. Then $\Psi \in L^p(X, \Sigma, \mu)$ and $\|\Psi_i - \Psi\|_p \to 0$.

This is the corollary which implies Eq. (5.15).

* Research supported in part by the National Science Foundation, Contract No. NSF-GP-8414.

[1] R. Dashen and D. H. Sharp, Phys. Rev. **165**, 1857 (1968).

[2] D. H. Sharp, Phys. Rev. **165**, 1867 (1968).

[3] C. Callan, R. Dashen, and D. H. Sharp, Phys. Rev. **165**, 1883 (1968).

[4] H. Sugawara, Phys. Rev. **170**, 1659 (1968), also investigates a theory in which the energy–momentum tensor is written in terms of local currents.

[5] G. Goldin, Ph.D. thesis, Princeton University, 1968.

[6] See, for example, H. Araki, J. Math. Phys. **1**, 492 (1960).

[7] S. Schweber, *An Introduction to Relativistic Quantum Field Theory* (Harper and Row, New York, 1961), p. 148.

[8] For an elementary discussion of distributions and their manipulation, the reader is referred to R. Streater and A. Wightman, *PCT, Spin and Statistics, and All That* (Benjamin, New York, 1964), Chap. 2.

[9] B. Renner, *Current Algebras and Their Applications* (Pergamon, New York, 1968), Chap. 8.

[10] See R. Haag and D. Kastler, J. Math. Phys. **5**, 848 (1964).

[11] J. Dixmier, *Les algèbres d'opérateurs dans l'espace hilbertien* (Gauthier-Villars, Paris, 1957), p. 17.

[12] K. Friedrichs, *Perturbation of Spectra in Hilbert Space*, 1960 Boulder, Colorado Lectures (American Mathematical Society, Providence, R.I., 1965), p. 41, provides an introduction to the Fock or particle number representation.

[13] The Fock representation of the anticommutation relations is also extensively discussed by R. Powers, Ph.D. thesis, Princeton University, 1967, Chap. I.

[14] S. Sternberg, *Lectures on Differential Geometry* (Prentice-Hall, Englewood Cliffs, N.J., 1964), p. 372, Appendix I, states the existence theorem for ordinary differential equations.

[15] For a discussion of vector fields and flows, see S. Lang, *Introduction to Differentiable Manifolds* (Interscience, New York, 1962), pp. 53–67. An important point is the fact that if $\mathbf{g}$ has components in $\mathfrak{Z}$, then $\varphi_t^{\mathbf{g}}(\mathbf{x})$ is defined *for all* t from $-\infty$ to $+\infty$. This fact follows from the corollary in Lang (p. 66) which states that a vector field on a compact manifold defines a flow for all t. The vector field $\mathbf{g}$ on $\mathbb{R}^s$ may be extended to the 1-point compactification of $\mathbb{R}^s$ by defining $g(\infty) = \mathbf{0}$.

[16] Stone's theorem may be found in F. Riesz and B. Sz.-Nagy, *Functional Analysis* (Ungar, New York, 1955), Leo F. Boron, transl., p. 383.

[17] E. Nelson, Ann. Math. **70**, 572 (1959), especially pp. 603–06. This article also provides a good discussion of analytic vectors.

[18] L. Schwartz, *Théorie des distributions* (Hermann, Paris, 1966).

[19] See L. Gårding, Proc. Natl. Acad. Sci. (U.S.) **33**, 331 (1947).

[20] Further discussion of the following theorem may be found in J. Grodnik and D. H. Sharp, Phys. Rev. D **1**, 1546 (1970).

[21] This follows from the fact that, for any set X in $\mathbb{R}^m$, $\mu^*(X) = \inf\{\mu(\mathcal{O}); X \subseteq \mathcal{O}, \mathcal{O} \text{ open}\}$, where μ^* denotes the outer measure. Cf. P. Halmos, *Measure Theory* (Van Nostrand, New York, 1950), p. 153.

[22] G. Goldin, J. Grodnik, R. Powers, and D. H. Sharp, to be published.

[23] J. Grodnik and D. H. Sharp, Phys. Rev. D **1**, 1531 (1970).

[24] The actual construction of the C^*-algebra of "uniformly almost periodic operators" is in J. Lew, Ph.D. thesis, Princeton University, 1960.

[25] Reference 11, especially Appendix I, Chaps. I.7 and II.6.

[26] Reference 21, p. 128.

[27] The use of the term "multiplier" to describe these χ's should not be confused with Mackey's use of the word; see, e.g., G. Mackey, Acta Math. **99**, 265 (1958).

[28] See Ref. 24.

[29] I. Gel'fand and N. Vilenkin, *Generalized Functions* (Academic, New York, 1964), Vol. 4, esp. Chaps. I and IV.

[30] A. Wightman, Rev. Mod. Phys. **34**, 845 (1962), Appendix IV.

[31] G. Mackey, Duke Math. J. **16**, 313 (1949).

[32] It is interesting to compare these remarks with the Mackey theory for (locally compact) semidirect products [G. Mackey, Ann. Math. **55**, 101 (1952), esp. p. 131]. There, the orbit structure is defined on the character group $\hat{N}$ of N, where $N \wedge G$ is the semidirect product. The representation determines a projection-valued measure P on $\hat{N}$. Under certain often verifiable conditions, the representation cannot be irreducible unless P is concentrated on a single orbit.

[33] See Refs. 20 and 23.

[34] The group $\mathfrak{W}$ may be suitably topologized (cf. Ref. 5).

[35] For the analogous axiom in the Wightman framework, see R. Jost, *The General Theory of Quantized Fields* (American Mathematical Society, Providence, R.I., 1965), p. 52.

[36] L. Schwartz, in *Proceedings of the International Congress of Mathematicians, 1950* (American Mathematical Society, Providence, R.I., 1952), Vol. I, p. 220.

[37] Since

$$\lim_{t \to \infty} \frac{(f \circ \varphi_t^{\mathbf{g}} - f)}{t} = \mathbf{g} \cdot \nabla f \text{ in } \mathfrak{S}.$$

[38] A. Dicke and G. Goldin, to be published.

[39] See also G. Goldin and D. H. Sharp, "Lie Algebras of Local Currents and their Representations," in the *1969 Battelle Rencontres: Group Representations*, V. Bargmann, Ed. (Springer-Verlag, Berlin, 1970), p. 300.

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 12, NUMBER 3 MARCH 1971

Eigenvalues of the Hill Equation to Any Order in the Adiabatic Limit*

JACQUES GUYARD, ANDRÉ NADEAU, GERMAIN BAUMANN, AND MARC R. FEIX
Groupe physique théorique et plasma, Université de Nancy, Nancy, France

(Received 13 July 1970)

To study the time-dependent linear oscillator, Lewis has recently introduced an auxiliary function w. One of the advantages of this function is that, in the adiabatic limit, a formal expansion of w in ϵ is possible (ϵ characterizing the slowness of the time variation). We show that, in this adiabatic limit, the eigenvalues of the Hill equation can be very easily deduced from w. Moreover, the computation of the ϵ^{2n}-order solution is much simpler if we use the Chandrasekhar method of higher invariants, which is shown to be equivalent to the Lewis expansion. Compact formulas, easy to handle on a computer, are obtained, and the method, which must be considered as a generalization to higher order of the WKB solution, is finally tested on the Mathieu equation.

I. INTRODUCTION

Recently, Lewis[1–3] has discovered a family of exact invariants for the equation of the time dependent linear oscillator

$$\ddot{q} + \Omega^2(t) \cdot q = 0. \tag{1}$$

In this paper we use these results to compute the eigenvalues of the Hill equation. [Equation (1), Ω being a periodic function of time]. Especially interesting is the case where Ω is a slowly varying function with t (adiabatic variation). Then, we can compute the eigenvalues to all orders in ϵ, where ϵ is a small parameter characterizing the slowness of the variation. We show that in this case the Lewis method is strictly equivalent to the iterative method proposed by Chandrasekhar[4] which gives, with a slight modification, a simple formula to compute the eigenvalues to high order in ϵ. This formula can be iterated on a computer using an algebraic language such as FORMAC. Finally, the method is tested on the Mathieu equation where results obtained by the lowest-order WKB method, the direct computation of the Hill determinant, and our method are compared.

II. APPLICATION OF THE LEWIS INVARIANT TO THE HILL EQUATION

To solve (1),an auxiliary function $w(t)$ is introduced, from which we define a new function Q and a new variable θ with

$$Q = q/w, \tag{2}$$

$$\theta = \int_0^t \frac{1}{w^2(t')}\, dt'. \tag{3}$$

Provided the $w(t)$ satisfy

$$\frac{d^2w}{dt^2} + \Omega^2(t)w = \frac{1}{w^3}, \tag{4}$$

(1) becomes

$$\frac{d^2Q}{d\theta^2} + Q = 0. \tag{5}$$

Although (4) is nothing but (1) plus a nonlinear term, the important point is that we can take for the solution of (4) any initial conditions.

For the Hill equation , Ω is a periodic function with period T. The solution of (1) is obtained from Floquet's theorem[5]

$$q = A \exp(i\mu t)F(t) + B \exp(-i\mu t)F(-t). \tag{6}$$

A and B are arbitrary constants and $F(t)$ is a periodic function of time with period T. μ is the eigenvalue of the Hill equation and is usually determined through the solution of the so called Hill determinant. When $\Omega(t)$ has many Fourier components we must use high rank determinant and we run into numerical difficulties.

Introducing the initial conditions $q\,(t = 0) = q_0$ and $\dot{q}\,(t = 0) = \dot{q}_0$ in (6) and taking into account the fact that $F(t)$ is periodic,we write the solution in a matrix form. After a little algebra we get:

$$\begin{bmatrix} q_T \\ \dot{q}_T \end{bmatrix} = \begin{bmatrix} \cos \mu T & i\dfrac{F_0}{\dot{F}_0 + i\mu F_0} \sin \mu T \\ i\dfrac{\dot{F}_0 + i\mu F_0}{F_0} \sin \mu T & \cos \mu T \end{bmatrix} \begin{bmatrix} q_0 \\ \dot{q}_0 \end{bmatrix} \tag{7}$$

with

$$F\,(t = 0) = F_0 \quad \text{and} \quad \left(\frac{dF}{dt}\right)_{t=0} = \dot{F}_0.$$

The eigenvalues λ of the matrix in (7) are solutions of the equation

$$\lambda^2 - 2\lambda \cos \mu T + 1 = 0.$$

Consequently,

$$\lambda = \exp(\pm i\mu T). \tag{8}$$

Another way to get q_T and $\dot{q}_T$ from q_0 and $\dot{q}_0$ is to use (2) and (3) to go from q and t to Q and θ and then to solve (5) and come back to q and t. Introducing

$$\Theta = \int_0^T \frac{1}{w^2(t)}\, dt, \tag{9}$$

we obtain

$$\begin{bmatrix} q_T \\ \dot{q}_T \end{bmatrix} = \begin{bmatrix} w_T & 0 \\ \dot{w}_T & w_T^{-1} \end{bmatrix} \begin{bmatrix} \cos\Theta & \sin\Theta \\ -\sin\Theta & \cos\Theta \end{bmatrix} \times \begin{bmatrix} w_0^{-1} & 0 \\ -\dot{w}_0 & w_0 \end{bmatrix} \begin{bmatrix} q_0 \\ \dot{q}_0 \end{bmatrix} = (A) \begin{bmatrix} q_0 \\ \dot{q}_0 \end{bmatrix}. \tag{10}$$

Computing μ is consequently equivalent to

(1) solving for $w(t)$ on any interval $[0, T]$ with any initial condition and computing $w(t)$, $\dot{w}(t)$, and Θ,

(2) looking for the eigenvalues of the matrix (A) as given by (10).

III. CASE OF SLOWLY VARYING $\Omega(t)$

Let us introduce explicitly the slow variation of Ω through a small parameter ϵ; Ω being supposed to be a function of the new variable $u = \epsilon t$: Equation (4) becomes

$$\epsilon^2 \frac{d^2w}{du^2} + \Omega^2(u)w = \frac{1}{w^3}. \tag{11}$$

We seek solutions of the form

$$w = w_0 + \epsilon^2 w_2 + \epsilon^4 w_4 + \cdots. \tag{12}$$

Introducing (12) in (11) and identifying the coefficients of ϵ^0, ϵ^2, etc., we obtain

$$\begin{cases} w_0 = \Omega^{-\frac{1}{2}}(u), \\ w_2 = \dfrac{1}{8}\,\Omega^{-\frac{7}{2}} \dfrac{d^2\Omega}{du^2} - \dfrac{3}{16}\,\Omega^{-\frac{9}{2}} \left(\dfrac{d\Omega}{du}\right)^2. \end{cases} \tag{13}$$

Notice that, in the solution of (11) as given by (13), we have no more the possibility to choose the initial conditions, which is all right since any initial conditions on w is permissible.

This approximation of the solution of (4) is now a periodic function of t with $w_T = w_0$ and $\dot{w}_T = \dot{w}_0$. The matrix (A) takes the simplified form

$$(A) = \begin{bmatrix} \cos\Theta - \sin\Theta w_0\dot{w}_0 & w_0^2 \sin\Theta \\ -\sin\Theta(w_0^{-2} + \dot{w}_0^2) & \cos\Theta + \sin\Theta w_0\dot{w}_0 \end{bmatrix}, \tag{14}$$

the eigenvalues of which are given by

$$\lambda^2 - 2\lambda\cos\Theta + 1 = 0. \tag{15}$$

The solutions are

$$\lambda = \exp(\pm i\Theta). \tag{16}$$

Comparing (16) and (8), we see that μ is given by

$$\mu = T^{-1}\Theta = T^{-1}\int_0^T \frac{dt}{w^2}, \tag{17}$$

which is the adiabatic solution (valid to all orders) for the eigenvalues of the Hill equation.

If we consider only the lowest-order solution for $w(t)$ given by

$$w_0 = \Omega^{-\frac{1}{2}},$$

we get

$$\mu = T^{-1}\int_0^T \Omega(t)\, dt, \tag{18}$$

which is nothing else but the well-known lowest-order WKB approximation.

IV. RESOLUTION OF THE LINEAR OSCILLATOR EQUATION BY CHANDRASEKHAR METHOD

To solve (1), Chandrasekhar[4] introduced a new variable t_1 and a new function q_1 with

$$\begin{aligned} \frac{dt_1}{dt} &= \Omega, \\ q_1 &= \Omega^{\frac{1}{2}} q. \end{aligned} \tag{19}$$

Equation (1) can be written

$$\frac{d^2q_1}{dt_1^2} + \Omega_1^2 q_1 = 0, \tag{20}$$

with

$$\Omega_1^2 = 1 + \Omega^{-\frac{3}{2}} \frac{d^2}{dt^2}\, \Omega^{-\frac{1}{2}}. \tag{21}$$

We introduce explicitly the Ω dependence through a variable $u = \epsilon t$. We get

$$\Omega_1^2 = 1 + \epsilon^2 \Omega^{-\frac{3}{2}} \frac{d^2}{du^2}\, \Omega^{-\frac{1}{2}}.$$

If Ω is a slowly varying function of t, then ϵ is small, and Ω_1 is equal to 1 plus an ϵ^2 order term.

Equation (20) is nothing but Eq. (1), where q, Ω, and t have been replaced by q_1, Ω_1, and t_1. We can repeat the operation introducing new variable and function t_2 and q_2 with

$$\begin{aligned} \frac{dt_2}{dt_1} &= \Omega_1, \\ q_2 &= \Omega_1^{\frac{1}{2}} q_1. \end{aligned} \tag{22}$$

We obtain

$$\begin{aligned} &\frac{d^2q_2}{dt_2^2} + \Omega_2^2 q_2 = 0, \\ &\Omega_2^2 = 1 + \Omega_1^{-\frac{3}{2}} \frac{d^2}{dt_1^2}\, \Omega_1^{-\frac{1}{2}}. \end{aligned} \tag{23}$$

Now Ω_2 is equal to 1 plus a term of order ϵ^4, and, if ϵ is small enough, the approximation $\Omega_2 = 1$ is an improvement on the preceding solution.

The process can be iterated, and Ω_n is now equal to 1 plus a term of order ϵ^{2n}.

Assuming that Ω_{n+1} is 1, we have to solve

$$\frac{d^2q_{n+1}}{dt_{n+1}^2} + q_{n+1} = 0$$

and to compute the new variable t_{n+1} and the different Ω_n with

$$\Omega_n^2 = 1 + \Omega_{n-1}^{-\frac{3}{2}} \frac{d^2}{dt_{n-1}^2} \Omega_{n-1}^{-\frac{1}{2}},$$

$$t_{n+1} = \int_0^t (\Omega\Omega_1 \cdots \Omega_n)\, dt'. \tag{24}$$

The advantage of this method is that it is an iterative one and is more suitable for a computational method.

V. EQUIVALENCE OF THE TWO METHODS IN THE ADIABATIC LIMIT

We introduce

$$\rho_n = (\Omega\Omega_1 \cdots \Omega_n)^{-\frac{1}{2}}. \tag{25}$$

Consequently,

$$\rho_{n+1}^{-4} = \rho_n^{-4}\Omega_{n+1}^2. \tag{26}$$

Taking (24) into account, we get

$$\rho_{n+1}^{-4} = \rho_n^{-4}\left(1 + \Omega_n^{-\frac{3}{2}} \frac{d^2}{dt_n^2} \Omega_n^{-\frac{1}{2}}\right) \tag{27}$$

but

$$\Omega_n^{-\frac{1}{2}} = \frac{\rho_n}{\rho_{n-1}}, \tag{28}$$

$$\frac{d}{dt_n}\Omega_n^{-\frac{1}{2}} = \frac{d}{dt}\left(\frac{\rho_n}{\rho_{n-1}}\right)\frac{dt}{dt_1}\frac{dt_1}{dt_2}\cdots\frac{dt_{n-1}}{dt_n},$$

$$\frac{d}{dt_n}\Omega_n^{-\frac{1}{2}} = \frac{d}{dt}\left(\frac{\rho_n}{\rho_{n-1}}\right)\rho_{n-1}^2 = \dot\rho_n\rho_{n-1} - \rho_n\dot\rho_{n-1},$$

$$\frac{d^2}{dt_n^2}\Omega_n^{-\frac{1}{2}} = \frac{d}{dt}(\dot\rho_n\rho_{n-1} - \rho_n\dot\rho_{n-1})\rho_{n-1}^2,$$

$$\frac{d^2}{dt_n^2}\Omega_n^{-\frac{1}{2}} = \ddot\rho_n\rho_{n-1}^3 - \rho_n\ddot\rho_{n-1}\rho_{n-1}^2. \tag{29}$$

Plotting (28) and (29) into (27), we get

$$\rho_{n+1}^{-4} = \rho_n^{-4}\left(1 + \frac{\rho_n^3}{\rho_{n-1}^3}(\ddot\rho_n\rho_{n-1}^3 - \rho_n\ddot\rho_{n-1}\rho_{n-1}^2)\right),$$

$$\rho_{n+1}^{-4} = \rho_n^{-4} + \frac{\ddot\rho_n}{\rho_n} - \frac{\ddot\rho_{n-1}}{\rho_{n-1}}.$$

This is a recurrence formula, so that

$$\rho_k^{-4} = \rho_{k-1}^{-4} + \frac{\ddot\rho_{k-1}}{\rho_{k-1}} - \frac{\ddot\rho_{k-2}}{\rho_{k-2}}. \tag{30}$$

By substituting each time the ρ_k^{-4}, $k = n, \cdots, 1$, in Eq. (30), we finally obtain

$$\rho_{n+1}^{-4} = \frac{\ddot\rho_n}{\rho_n} + \rho_1^{-4} - \frac{\ddot\rho_0}{\rho_0}. \tag{31}$$

But

$$\rho_1^{-4} = \rho_0^{-4}\Omega_1^2 = \rho_0^{-4} + \rho_0^{-1}\ddot\rho_0. \tag{32}$$

Consequently, taking into account (32) and $\rho_0 = \Omega^{-\frac{1}{2}}$, we see that the recurrence formula between ρ_n and ρ_{n+1} is therefore

$$\rho_{n+1}^{-4} = \ddot\rho_n/\rho_n + \Omega^2 \quad \text{with} \quad \rho_0 = \Omega^{-\frac{1}{2}}. \tag{33}$$

If we notice that we need only ρ_n (and not Ω, $\Omega_1, \cdots, \Omega_n$), we see that (33) gives in a very compact way the solution of the Chandrasekhar high-order invariants; especially interesting is the fact that we get rid of all the intermediate changes of variables. Moreover, (33) strongly reminds us of the Lewis equation [Eq. (4)]. To solve it, we have to compute the series

$$v_n = w_0 + \epsilon^2 w_2 + \cdots + \epsilon^{2n} w_{2n}. \tag{34}$$

In order to get the next term v_{n+1}, we introduce (34) in

$$\epsilon^2 \frac{1}{w}\frac{d^2w}{du^2} + \Omega^2 = w^{-4}.$$

The expansion including the term of order ϵ^{2n+2} of the left member is obtained by replacing w by its v_n approximation. Consequently, in the right member, we must use the v_{n+1} expansion; of course, the two terms will have the same expansion up to ϵ^{2n+2} (included). We have

$$\epsilon^2 \frac{1}{v_n}\frac{d^2v_n}{du^2} + \Omega^2 = v_{n+1}^{-4} + A\epsilon^{2n+4}. \tag{35}$$

Therefore, we see that, if we expand the Chandrasekhar ρ_n up to order ϵ^{2n}, we obtain v_n. Then, ρ_n and v_n have the same expansion in ϵ. Moreover, if the series are convergent, ρ_∞ obeys (4) and, strictly, is equal to w.

Now it must be pointed out that, from a practical computational point of view, the two methods are not completely equivalent. Indeed, if results are desired that are correct to order ϵ^{2n} and that contain no higher-order corrections, then both the quantities in Eqs. (33) and (35) have to be expanded in powers of ϵ which is a tedious analytical work. But, in fact, we can use results correct to order ϵ^{2n} containing some part of the higher-order corrections. Then Eq. (33) directly solves this problem and so does Eq. (35), where we just ignore the last term $A\epsilon^{2n+4}$.

Now we have just to carry successive derivations and an algebraic language (as FORMAC) can be used.

Working on power expansion of Eq. (4), Lewis[3] has also used FORMAC to compute $v_1, \cdots, v_3$. Here we get $\rho_0, \cdots, \rho_6$, corresponding to $v_0, \cdots, v_6$, by using directly Eq. (33).

Details on this computation will be published elsewhere.

To come back to the Hill's equation we see that in the adiabatic limit the eigenvalues up to order ϵ^{2n} (included) are given by

$$\mu_n = T^{-1}\int_0^T \rho_n^{-2}\, dt,$$

with ρ_n given by (33).

VI. APPLICATION: COMPUTATION OF THE EIGENVALUES OF MATHIEU'S EQUATION

To test the method, we take $\Omega^2 = A - 2Q\cos 2t$.

In this case (see Ref. 5) the eigenvalues are a solution of

$$\Delta(0)\sin^2(\tfrac{1}{2}\pi\sqrt{A}) = \sin^2(\tfrac{1}{2}\pi\mu),$$

where $\Delta(0)$ is the following determinant:

$$\begin{vmatrix}
\cdot & \cdot & \cdot & \cdot & \cdot & \\
\cdot\ 1 & \frac{Q}{4^2-A} & 0 & 0 & 0 & \cdot \\
\cdot\ \frac{Q}{2^2-A} & 1 & \frac{Q}{2^2-A} & 0 & 0 & \cdot \\
\cdot\ 0 & \frac{Q}{0^2-A} & 1 & \frac{Q}{0^2-A} & 0 & \cdot \\
\cdot\ 0 & 0 & \frac{Q}{2^2-A} & 1 & \frac{Q}{2^2-A} & \cdot \\
\cdot\ 0 & 0 & 0 & \frac{Q}{4^2-A} & 1 & \cdot \\
\cdot & \cdot & \cdot & \cdot & \cdot &
\end{vmatrix}$$

$\Delta(0)$, because of its very simple form, can be computed to very high order by recurrent formulas avoiding the round-off errors which occur with classical methods.

The results shown on the Figs. 1 and 2 are given for $2Q/A = 0.4$ and for values of A going from 10 to 1000. The adiabatic approximation corresponds to large values of A.

We compare three results:

(1) The lowest-order WKB solution: In this approximation the normalized eigenvalue $\mu/\sqrt{A}$ is a constant independent of A and equal, for $2Q/A = 0.4$, to 0.98959875.

(2) The adiabatic solution μ_1 and μ_2 obtained, respectively, with ρ_1 and ρ_2.

(3) The results μ of the direct solution of the Hill determinant. For A greater than 20 ($A^{-\frac{1}{2}} < 0.224$), Fig. 1 shows μ_1 and μ (μ_2 cannot be distinguished from μ_1). The agreement is excellent; for A smaller than 40 ($A^{-\frac{1}{2}} > 0.158$), Fig. 2 shows μ_1, μ_2, and μ.

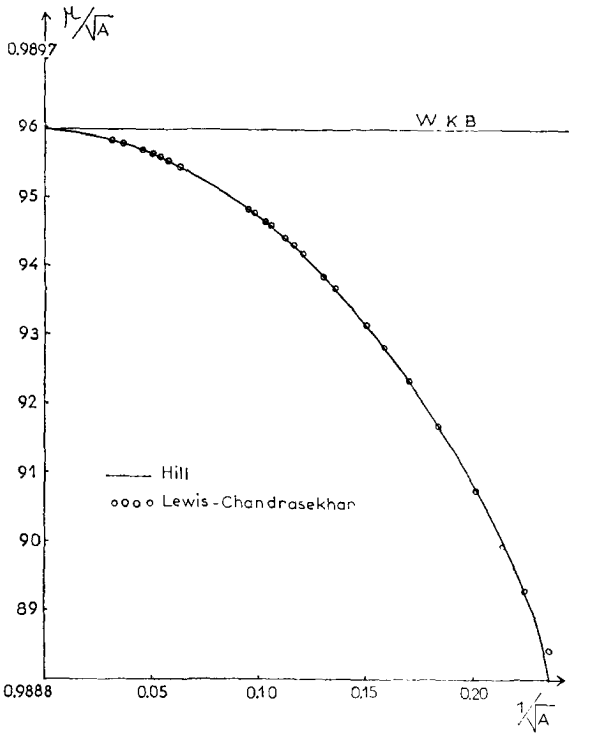

FIG. 1.

An important consequence of the adiabatic approximation is shown on this last figure, namely the disappearance of the unstable zones of the Mathieu

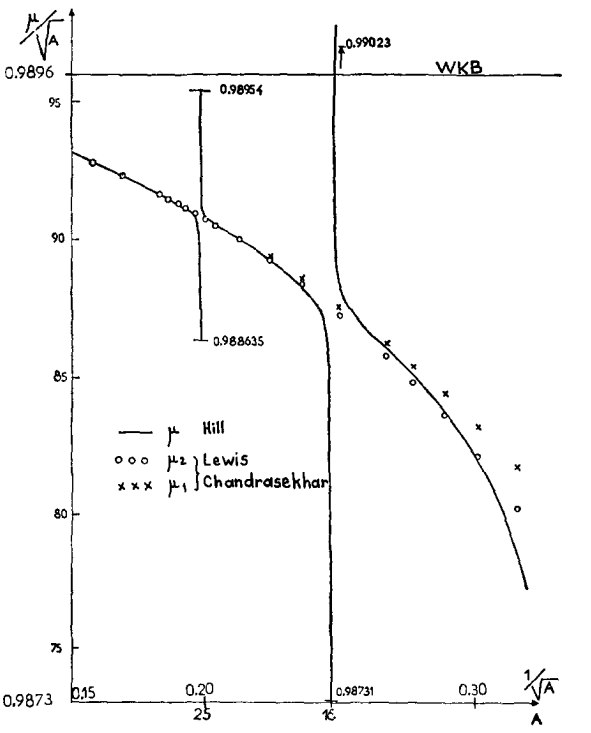

FIG. 2.

equation. In fact, this phenomena occurs for all values of A closed to the square of an integer, but for $A > 36$ the unstable zones are so small that they cannot be shown on Fig. 1. If Q/A is small enough the width of each zone is of the order of $A^{-\frac{3}{2}}(Q/A)^{\sqrt{A}}$ and in the adiabatic limit such an effect is totally ignored.

On the μ curve of Fig. 2 these instable zones appear around $A = 25$ and $A = 16$. We notice that μ_2 just bridges the discontinuities of the exact μ curve and comes back on this last curve while the μ_1 approximation is definitively too large. Finally, for the value $A < 10$ the adiabatic limit has no meaning.

We notice the large improvement of the μ_1 approximation upon the lowest-order WKB method.

ACKNOWLEDGMENTS

Interesting discussion on the problem and decisive help on FORMAC program from Dr. J. Bitoun are gratefully acknowledged. Also, the authors thank the referee for a useful remark.

* A part of this work was done during the tenure of a NAS-NRC Senior Research Associateship at the Goddard Space Flight Center, Greenbelt, Maryland.

[1] H. R. Lewis, Phys. Rev. Letters **18**, 510 (1967).
[2] H. R. Lewis, Phys. Rev. **172**, 1313 (1968).
[3] H. R. Lewis, J. Math. Phys. **9**, 1976 (1968).
[4] S. Chandrasekhar, *The Plasma in a Magnetic Field*, R. K. M. Landshoff, Ed. (Stanford U.P., Palo Alto, Calif., 1958).
[5] E. T. Whittaker and G. N. Watson, *A Course of Modern Analysis* (Cambridge U.P., Cambridge, 1966).

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 12, NUMBER 3 MARCH 1971

The Reduction $O(3, 1) \supset O(2, 1) \supset O(1, 1)$

N. W. MACFADYEN
Carnegie-Mellon University, Pittsburgh, Pennsylvania 15213

(Received 26 June 1970)

We set up a representation of the principal series of the group $SL(2, C)$ in terms of the sequence of noncompact subgroups $SL(2, C) \supset SU(1, 1) \supset O(1, 1)$. The basis functions are just the cross-basis matrix elements of $SU(1, 1)$ between $O(2)$ and $O(1, 1)$ bases, and we derive much material on these before using them to calculate the representation functions. The latter have two pairs of discrete labels distinguishing between equivalent representations occurring in the reduction, and these are related to two discrete reflection operators that are introduced in order to obtain a maximal Abelian set.

INTRODUCTION

In a previous paper[1] (to be referred to as I) we examined the reductions $SL(2, C) \supset SU(1, 1) \supset O(2)$ and $SL(2, R) \supset O(1, 1)$ and gave a complete account of the global properties of the irreducible representations in the corresponding bases. We did not, however, consider the reduction in the sequence of noncompact subgroups $SL(2, C) \supset SU(1, 1) \supset O(1, 1)$; while this has no very new features, it is of interest to see how our previous results combine, and so here we present the theory of this reduction for the principal continuous series of representations.

The interesting part of the problem is concerned with the basis functions we must use. There are now no fewer than four sets of these, distinguished by a pair of degeneracy labels, because each group representation contains those of its subgroup twice over; and they are just the "cross-basis" matrix elements of $SU(1, 1)$ that we can label schematically

$$\langle O(2)| \exp(iK_2\xi) |O(1, 1)\rangle.$$

Much of the paper is devoted to defining and examining these functions; indeed the first two sections and Appendix A are devoted entirely to these preliminaries, and only in Sec. III do we at last discuss the eponymous problem.

We achieve a formulation of the reduction that in principle is complete, but in practice the only coset class whose representation functions we can calculate is that one which we were unable to treat in I. As expected, the asymptotic behavior of these functions in the complex plane of the Casimir operator σ is just that of I; we can also show that with this basis we do not obtain matrix elements of an element $a \notin SU(1, 1)$ that are simultaneously of simple behavior in both σ and j, thus disposing of any hopes of a "completely second-kind" set of representation functions. Hence, the main interest (or usefulness!) of the work lies in the cross-basis matrix elements; there are some indications that these may be of use in the multi-Regge theory.

This paper is essentially a sequel to I, to which we refer constantly; hence all the results of that paper

equation. In fact, this phenomena occurs for all values of A closed to the square of an integer, but for $A > 36$ the unstable zones are so small that they cannot be shown on Fig. 1. If Q/A is small enough the width of each zone is of the order of $A^{-\frac{3}{2}}(Q/A)^{\sqrt{A}}$ and in the adiabatic limit such an effect is totally ignored.

On the μ curve of Fig. 2 these instable zones appear around $A = 25$ and $A = 16$. We notice that μ_2 just bridges the discontinuities of the exact μ curve and comes back on this last curve while the μ_1 approximation is definitively too large. Finally, for the value $A < 10$ the adiabatic limit has no meaning.

We notice the large improvement of the μ_1 approximation upon the lowest-order WKB method.

ACKNOWLEDGMENTS

Interesting discussion on the problem and decisive help on FORMAC program from Dr. J. Bitoun are gratefully acknowledged. Also, the authors thank the referee for a useful remark.

* A part of this work was done during the tenure of a NAS-NRC Senior Research Associateship at the Goddard Space Flight Center, Greenbelt, Maryland.

[1] H. R. Lewis, Phys. Rev. Letters **18**, 510 (1967).
[2] H. R. Lewis, Phys. Rev. **172**, 1313 (1968).
[3] H. R. Lewis, J. Math. Phys. **9**, 1976 (1968).
[4] S. Chandrasekhar, *The Plasma in a Magnetic Field*, R. K. M. Landshoff, Ed. (Stanford U.P., Palo Alto, Calif., 1958).
[5] E. T. Whittaker and G. N. Watson, *A Course of Modern Analysis* (Cambridge U.P., Cambridge, 1966).

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 12, NUMBER 3 MARCH 1971

The Reduction $O(3, 1) \supset O(2, 1) \supset O(1, 1)$

N. W. MACFADYEN
Carnegie-Mellon University, Pittsburgh, Pennsylvania 15213

(Received 26 June 1970)

We set up a representation of the principal series of the group $SL(2, C)$ in terms of the sequence of noncompact subgroups $SL(2, C) \supset SU(1, 1) \supset O(1, 1)$. The basis functions are just the cross-basis matrix elements of $SU(1, 1)$ between $O(2)$ and $O(1, 1)$ bases, and we derive much material on these before using them to calculate the representation functions. The latter have two pairs of discrete labels distinguishing between equivalent representations occurring in the reduction, and these are related to two discrete reflection operators that are introduced in order to obtain a maximal Abelian set.

INTRODUCTION

In a previous paper[1] (to be referred to as I) we examined the reductions $SL(2, C) \supset SU(1, 1) \supset O(2)$ and $SL(2, R) \supset O(1, 1)$ and gave a complete account of the global properties of the irreducible representations in the corresponding bases. We did not, however, consider the reduction in the sequence of noncompact subgroups $SL(2, C) \supset SU(1, 1) \supset O(1, 1)$; while this has no very new features, it is of interest to see how our previous results combine, and so here we present the theory of this reduction for the principal continuous series of representations.

The interesting part of the problem is concerned with the basis functions we must use. There are now no fewer than four sets of these, distinguished by a pair of degeneracy labels, because each group representation contains those of its subgroup twice over; and they are just the "cross-basis" matrix elements of $SU(1, 1)$ that we can label schematically

$$\langle O(2)| \exp(iK_2\xi) |O(1, 1)\rangle.$$

Much of the paper is devoted to defining and examining these functions; indeed the first two sections and Appendix A are devoted entirely to these preliminaries, and only in Sec. III do we at last discuss the eponymous problem.

We achieve a formulation of the reduction that in principle is complete, but in practice the only coset class whose representation functions we can calculate is that one which we were unable to treat in I. As expected, the asymptotic behavior of these functions in the complex plane of the Casimir operator σ is just that of I; we can also show that with this basis we do not obtain matrix elements of an element $a \notin SU(1, 1)$ that are simultaneously of simple behavior in both σ and j, thus disposing of any hopes of a "completely second-kind" set of representation functions. Hence, the main interest (or usefulness!) of the work lies in the cross-basis matrix elements; there are some indications that these may be of use in the multi-Regge theory.

This paper is essentially a sequel to I, to which we refer constantly; hence all the results of that paper

which we need are assumed, and not further discussed here. The notation and other conventions are to be found there too.

I. ISOMORPHISMS OF THE CARRIER SPACE OF $SL(2, R)$

We shall wish later to expand arbitrary functions on $SU(1, 1)$ in terms of the cross-basis matrix elements. In this section we investigate the precise meaning of these, and in the following show that they have all the properties we need; but because it is rather clearer to work with $SL(2, R)$, we shall carry out our calculations with that group and make use of the isomorphism between the two to transform our results only at a later stage.

Consider then the carrier space D_j of a representation j of $SL(2, R)$; we can realize this in three different ways of interest (I, Sec. 1.1)—as functions defined over the real line (parametrized by x), the unit circle (α), or a two-sheeted hyperbola (β_τ). The isomorphisms

$$u: D_j(x) \to D_j(\alpha),$$
$$v: D_j(x) \to D_j(\beta) = \bigcup_\tau D_j^\tau(\beta) \tag{1}$$

are defined by the relations, valid for any $f \in D_j(x)$,

$$(u:f)(\alpha) = \lambda^{-2j} f(x), \quad x = k\alpha,$$
$$(v:f)^\tau(\beta) = \lambda^{-2j} f(x), \quad x = k\epsilon^\rho\beta, \tag{1'}$$

where we have introduced the notation $\lambda^{2j} = |\lambda|^{2j}(\operatorname{sgn}\lambda)^\omega$. From these we define $w = vu^{-1}$, finding

$$(w^{-1}: \phi^\tau)(\alpha) = \lambda^{-2j}\phi^\tau(\beta), \quad \epsilon^\rho\beta = k\alpha, \tag{2}$$

for any $\phi^\tau \in D_j^\tau(\beta)$. If the equation for β has no solution for a given α, we define the right-hand side of (2) to be zero.

These mappings (which are isometries with the measures $d\alpha$, $d\beta$, and $2\,dx$) can be extended to the Hilbert spaces $\mathcal{H}_j$, and therefore specify equivalences between UIR's in each space; we can illustrate them by the commutative diagram of Fig. 1, where the notation is self-explanatory.

Now define the map $S_\tau^j: D_j(\beta) \to D_j(\alpha)$ by

$$S_\tau^j = T_\tau^j(\alpha)w^{-1} = w^{-1}T_\tau^j(\beta), \tag{3}$$

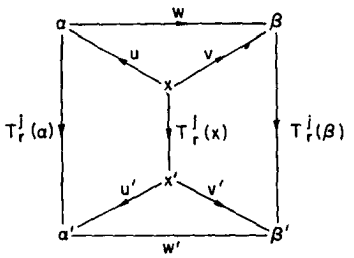

FIG. 1. The equivalences u, v, and w between the spaces $D_j(x)$, $D_j(\alpha)$, and $D_j(\beta)$.

where the labels in parentheses indicate the space in which T acts. Then we can choose bases $\{\psi_m\}$ and $\{\phi_\mu^\tau\}$ of $D_j(\alpha)$ and $D_j^\tau(\beta)$ (or, more strictly, of their duals) and define the function

$$D_{m\mu}^{j\tau}(r) = (\psi_m, S_\tau^j: \phi_\mu^\tau)$$
$$= \int d\alpha\,\overline{\psi_m(\alpha)}\lambda^{2j}\phi_\mu^\tau(\beta), \quad \alpha r = k\epsilon^\rho\beta. \tag{4}$$

We shall refer to this as the cross-basis matrix element. It is easy to see that under the group it transforms as

$$D_{m\mu}^{j\tau}(\alpha\xi\beta) = e^{im\alpha}D_{m\mu}^{j\tau}(\xi)e^{i\mu\beta}, \tag{5}$$

$$D_{m\mu}^{j\tau}(rr') = \sum_{m'} D_{mm'}^{j}(r)D_{m'\mu}^{j\tau}(r')$$
$$= \sum_{\tau'}\int d\mu' D_{m\mu'}^{j\tau'}(r)D_{\mu'\mu}^{j\tau'\tau}(r'); \tag{6}$$

the function itself is calculated in the Appendix, and we find

$$D_{m\mu}^{j+}(\xi) = e^{-\frac{1}{2}i\pi(j+1)}D_{im,\mu}^{j++}(\xi - \tfrac{1}{2}i\pi), \tag{7}$$

$$D_{m\mu}^{j-}(\xi) = e^{-im\pi}D_{m,-\mu}^{j+}(-\xi), \tag{8}$$

where $\xi = \exp(iK_2\xi)$ and the function on the right of (7) is that discussed in I. These relations must be treated with care because of the cut structure.

II. EXPANSION THEOREMS ON $SL(2, R)$

We wish to derive orthogonality and completeness relations for the cross-basis matrix elements: To do so, we must make use of basis-independent Fourier transforms[2,3] over the group. Recall that the Fourier Transform (FT) of an indefinitely differentiable function $f(r)$ of compact support is the operator

$$T_f^j = \int f(r)T_r^j\,d\mu(r).$$

We investigate here the operator $S_f^j = T_f^j w^{-1}$, and find

$$(S_f^j: \phi^\tau)(\alpha) = \int f(r)\lambda^{2j}\phi^\tau(\beta)\,d\mu(r), \tag{9}$$

where $\alpha r = k\epsilon^\rho\beta$. Once again, α and τ are fixed, so the integral is over only that part of $SL(2, R)$ where the parametrization exists. This can be written as

$$\int \tilde{f}(\alpha, \beta, \tau \mid j)\phi^\tau(\beta)\,d\beta,$$

where

$$\tilde{f}(\alpha, \beta, \tau \mid j) = \int f(\alpha^{-1}k\epsilon^\rho\beta)\lambda^{2j}\,d\mu_t(k), \tag{10}$$

and we have set $d\mu(r) = d\beta\,d\mu_t(k)$. We shall often refer to the kernel $\tilde{f}$ as the (cross-basis) FT. Now suppose that $f(r)$ satisfies the condition

$$f(\alpha r\beta) = e^{-im\alpha}f(r)e^{-i\mu\beta}; \tag{11}$$

we can certainly decompose any $f \in C^\infty(r)$ into such $f_{m\mu}$ by a classical Fourier transform in each one-parameter subgroup, and we shall assume it done. Then we find

$$\tilde{f}(\alpha, \beta, \tau \mid j)_{m\mu} = \psi_m(\alpha)\overline{\phi^\tau_\mu(\beta)}\tilde{f}_{m\mu}(\tau, j), \tag{12}$$

where

$$\tilde{f}_{m\mu}(\tau, j) = \int f_{m\mu}(k\epsilon^\rho)\lambda^{2j}\, d\mu_l(k).$$

Upon setting $k\epsilon^\rho = \alpha\xi\beta^{-1}$ and $d\mu_l(k) = d\alpha\, d\omega(\xi)$, this becomes at last

$$\tilde{f}_{m\mu}(\tau, j) = \int d\mu(r) f(r) D^{j\tau}_{m\mu}(r), \tag{13}$$

and a trivial modification of Gel'fand's treatment[2,3] tells us that the inversion formula is

$$f(r) = \sum_{m,\tau} \int d\mu \int dM(j) \tilde{f}_{m\mu}(\tau, j)\overline{D^{j\tau}_{m\mu}(r)}. \tag{14}$$

Combining these two relations, we find at last the completeness and orthogonality formulas we have been seeking:

$$\int d\mu(r) D^{j\tau}_{m\mu}(r)\overline{D^{j'\tau'}_{m'\mu'}(r)} = \delta_{mm'}\delta_{\tau\tau'}\delta(\mu - \mu')\delta_{jj'}(j - j'),$$

$$\sum_{m,\tau} \int d\mu \int dM(j) D^{j\tau}_{m\mu}(r)\overline{D^{j\tau}_{m\mu}(r')} = \delta_\mu(rr'^{-1}). \tag{15}$$

The measure $d\mu(j)$ is that of [I, (2.8)]:

$$\int f_m(j)\, dM(j) = \sum_{k>0}^{|m|-1} (2k+1) f_m(k) + \frac{1}{2i}\int_{-\frac{1}{2}-i\infty}^{-\frac{1}{2}+i\infty} dj(2j+1)\cot \pi(j-m) f_m(j),$$

and the invariant integral over the group is

$$\int d\mu(r) = (4\pi)^{-2}\int_0^{4\pi} d\alpha \int_{-\infty}^{\infty} d\beta \int_{-\infty}^{\infty} d(\sinh \xi) \tag{16}$$

under the obvious parametrization $r = \alpha\xi\beta$. The δ functions in (15) are of course with respect to these measures.

Notice that the completeness relation requires a sum over τ-labels. This is exactly analogous to the expansion formulas[4] for functions defined on a hyperboloid parametrized by a hyperbolic coordinate system, which are well known; and of course our expressions reduce to those in the spinless case $m = 0$. It is worth remarking that (15) can be derived entirely independently by making use of Popov's method[5,6] of generalizing the Gel'fand–Graev transform.[2,4]

III. THE DECOMPOSITION

We are now in a position to approach the main problem of this paper; but first we must say a few words about notation. It was convenient in I to use the same symbols for the corresponding quantities associated with $SL(2, C)$ and $SL(2, R)$, and it caused no confusion there; and since with the derivation of the expansion formulas (15) we have concluded our treatment of $SL(2, R)$ as such, we shall do the same here. Henceforth, the letters ρ and τ will refer to $SL(2, C)$, while the same quantities associated with the subgroup will be denoted r and t; the matrix k will be complex. We shall in fact be considering the reduction $SL(2, C) \supset SU(1, 1) \supset O(1, 1)$; the group V is derived from R by the mapping

$$M: r \in R = M^\dagger r M \in V; \quad M = \frac{1}{\sqrt{2}}\begin{pmatrix} 1 & i \\ i & 1 \end{pmatrix},$$

but the details of this will not concern us.

Recall then[1] that we established a representation $\chi = (j_0, \sigma)$ of $SL(2, C)$ by the operators

$$T^\chi_a: f^\tau(v) = \lambda^\chi f^{\tau'}(v'), \tag{17}$$

$$\epsilon^\rho v a = k\epsilon^{\rho'} v', \tag{18}$$

where $a \in C$, $v \in V$, and $f \in D_\chi$. We now ask that the operator T^χ_β representing a boost of magnitude β along the 1 axis, be diagonal; in conjunction with the covariance condition on f arising from the ambiguity in phase of λ, this implies

$$f^\tau(\alpha v \beta) = e^{i\tau j_0 \alpha} f^\tau(v) e^{i\mu\beta}, \tag{19}$$

where μ is the eigenvalue of K_1. As we anticipated, we need to expand $f^\tau(v)$ in terms of the cross-basis matrix elements:

$$f^\tau(v) = \sum_\tau \int d\mu\, dM(j) \tilde{f}^{j\tau}_{\mu t} D^{jt}_{\tau j_0, \mu}(v),$$

$$\tilde{f}^{j\tau}_{\mu t} = \int d\mu(v) f^\tau(v)\overline{D^{jt}_{\tau j_0, \mu}(v)}; \tag{20}$$

we are assured that this is sufficient because $C^\infty(v)$ is dense in D^τ_χ.

We therefore have no fewer than four independent functions f^τ_t, and it is most desirable to give these some interpretation. As we saw in I, the decomposition $C \supset V$ corresponds to taking a section of the hyperboloid $H_3 = \{x \mid x_0^2 - x^2 = 1\}$ by a plane at $x_3 =$ const, so that τ specifies the sign of x_0; if then we cut this by a further plane at $x_2 =$ const (Fig. 2), we are left with two hyperbolas H_1^τ whose τ-labels have the identical meaning, so that we might suppose that only a single pair of labels was required.

However, to specify the upper-sheet values of a function defined over H_2 and transforming under a given UIR of $SL(2, R)$ or $SU(1, 1)$ (that is, one whose Gel'fand transform[2] is a homogeneous function on the cone), it is not sufficient[4] to give its values upon H_1^+: We need also those on the hyperbola $\bar{H}_1^+$, which is the

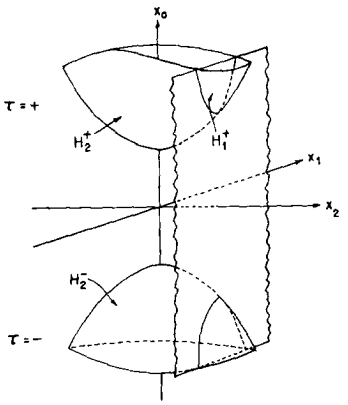

FIG. 2. The cross section $x_3 =$ constant of the hyperboloid $x_0^2 - x^2 = 1$ cut by the plane $x_2 =$ const.

intersection of H_2^+ with the plane $x_2 = -$const—or, equivalently, those on its reflection H_1^-. But in the reduction $O(3, 1) \supset O(2, 1) \supset O(1, 1)$, with which we are here concerned, we require an arbitrary function on H_2—that is, one whose values on the two sheets are independent—because both sheets are needed to specify the function on H_3^+: And hence we indeed need it on all four hyperbolas H_1^τ, $\bar{H}_1^\tau$. We shall label the four functions by $f_t^\tau(\beta)$, where τ denotes (sgn x_0) and t is (sgn x_2).

An equivalent explanation, but with an algebraic rather than a geometric basis, is afforded by noticing that the operators corresponding to $\{\sigma, j_0, j, \mu\}$ do not form a maximal Abelian set: They can be augmented by a pair of reflections T and R that do not both belong to $SL(2, C)$. With the standard homomorphism $SL(2, C) \to O(3, 1)$, we can write these as

$$T:(x_0, \mathbf{x}) = (-x_0, -x_1, -x_2, x_3),$$
$$R:(x_0, \mathbf{x}) = (x_0, x_1, -x_2, -x_3),$$

and it is easy to check that except for $T: j_0 \to -j_0$ these commute with the former operators. Clearly, T has the significance of a time reflection followed by a rotation through π about the 3 axis; R is just a similar rotation about the 1 axis. Adjoining these operators to the previous ones gives us the complete set $\{\sigma, |j_0|, j, \mu, T, R\}$. Our basis vectors are not the eigenfunctions of T and R, but rather those linear combinations of them with definite signs of x_0 and x_2.

It is now easy to see how the discrete series of $SU(1, 1)$ enters the picture. Recall the results of I: We know that upon H_2^+ only the positive discrete series k^+ is needed to expand $f(v)$ (in addition, of course, to the principal continuous series of representations), and so we find that the Fourier transform $\tilde{f}_{\mu t}^{k^+}$ of $f_t^{k^+}(\beta)$ vanishes on the half-line $\mu t < 0$. Similarly, the lower sheet H_2^- requires only the negative series k^-, and $\tilde{f}_{\mu t}^{k^-}$ vanishes for $\mu t > 0$.

Let us return to the representation. If we expand $f^\tau(v)$ by (20), then the presence of two terms in the completeness relation means that we need to calculate a set of integral kernels labeled by all four parameters τ, τ', t, t':

$$D^{\chi\tau\tau'}_{j\mu t, j'\mu' t'}(a) = (\phi^{j\tau}_{\mu t}, T^\chi_a \phi^{j'\tau'}_{\mu' t'}) = \int \overline{D^{jt}_{\tau j_0 \mu}(v)} \lambda^\chi D^{j't'}_{\tau' j_0 \mu'}(v')\, d\mu(v), \quad (21)$$

where

$$\epsilon^{\rho'} v a = k \epsilon^\rho v' \quad (22)$$

defines, for fixed ρ and ρ', both the variable v' and the range of integration. In the special case $a = v_0 \in V$, it is easy to see using (6) and (15) that this reduces to

$$D^{\chi\tau\tau'}_{j\mu t, j'\mu' t'}(v_0) = \delta_{\tau\tau'} \delta_M(j - j') D^{jtt'}_{\mu\mu'}(v_0) \quad (23)$$

as required, and so we need only calculate new functions for representatives of the three double coset classes[1] (C1)–(C3).

Unfortunately, the only one of these that it is feasible to approach is (C3): For if we choose as representative of this the element θ',

$$\theta' = \begin{pmatrix} \cos \tfrac{1}{2}\theta & i \sin \tfrac{1}{2}\theta \\ i \sin \tfrac{1}{2}\theta & \cos \tfrac{1}{2}\theta \end{pmatrix} = \exp(i\theta J_1), \quad (24)$$

we see that it commutes with β and so the μ-labels will be unchanged—that is, (21) will contain a factor $\delta(\mu - \mu')$. Both of the other classes will change β and so effectively prevent us from calculating the integral in any useful form.

In Appendix B we have calculated the matrix element for the choice $\tau = \tau' = t = t' = +$; the final expression is found in (B5), and is superficially quite similar to the expressions given in I for $d^{\chi\tau\tau'}_{lmj}(\xi)$. From this and (B8) we can obtain the remaining matrix elements of $0 < \theta < \pi/2$ by using the identities

$$D^{\chi\tau\tau'}_{l\mu t, j\mu t'}(\theta') = e^{2\pi i j_0(\rho+\rho')} D^{\chi-\tau-\tau'}_{l\mu-t, j\mu-t'}(\theta'),$$
$$D^{\chi++}_{l\mu t, j\mu t'}(\theta') = e^{2\pi i j_0(\tau+\tau')} \overline{D^{\chi++}_{j-\mu-t', l-\mu-t}(\theta')} \quad (25)$$

which we obtain by manipulation of the defining integrals, and the curious intertwining relation derived from (A12):

$$e^{2\pi i j_0} \alpha'^l_\mu D^{\chi+\tau}_{l\mu-, j\mu' t}(a) = a^l_m D^{\chi+\tau}_{l\mu+, j\mu' t}(a) - \alpha^l_\mu D^{\chi+\tau}_{l\mu+, j\mu' t}(a), \quad (26)$$

which is valid for all $a \in SL(2, C)$.

Now these results apply in the first instance only to the region $0 < \theta < \pi/2$ because of considerations of the phase of λ in the parametrization (22). To extend them to other regions we use the discrete transforms

$$D^{\chi\tau\tau'}_{l\mu t, j\mu t'}(\theta' = \pi) = e^{2\pi i \rho j_0} \delta(\mu - \mu') \delta_M(l - j) \delta_{t,-t'} \delta_{\tau,-\tau'},$$
$$D^{\chi\tau\tau'}_{l\mu t, j\mu t'}(\epsilon) = e^{2\pi i \rho \tau j_0} \delta(\mu + \mu') \delta_M(l - j) \delta_{tt'} \delta_{\tau,-\tau'} \quad (27)$$

and the two fundamental relations of unitarity and representation property:

$$D^{\chi\tau\tau'}_{j\mu t,j'\mu' t'}(a) = \overline{D^{\chi\tau'\tau}_{j'\mu' t',j\mu t}}(a^{-1}), \tag{28}$$

$$D^{\chi\tau\tau'}_{j\mu t,j'\mu' t'}(aa') = \sum_{\tau'',t''}\int d\mu''\, dM(l) D^{\chi\tau\tau''}_{j\mu t,l\mu'' t''}(a) D^{\chi\tau''\tau'}_{l\mu'' t'',j\mu t}(a'). \tag{29}$$

By applying these we can obtain all the matrix elements of the class (C3).

Finally, let us consider the behavior of the ME's in the complex planes of l and σ. Since the choice of basis cannot affect the decomposition $C \supset V$, we know that the σ-behavior is just that found in I; hence we do not obtain thus a "second-kind" decomposition of $D^{\chi}(\theta')$, as indeed is easily seen from (B3). We might, however, have hoped that all these matrix elements would be of simple behavior in l and j; but because the behavior of the basis functions depends upon the sign of ξ [see the remarks after (A9)], any integral like (B3), which includes in its range the point $\xi = 0$, will not have simple behavior. Since in general the defining integrals do contain this point, we do not find matrix elements of $a \notin V$ that are simultaneously second-kind in l, j, and σ.

This then concludes our description of those problems with relatively simple solutions. Curiously enough, the completeness relations for the expansion of a function defined over a $3+1$ hyperboloid parametrized by a Lobachevskii coordinate system have never been given, although the basis functions may be found in Ref. 4; the authors there considered that the absence of an azimuthal angle destroyed the usefulness of the system. Certainly the complications introduced by this reduction of the group lower our faith in its applicability.

ACKNOWLEDGMENT

I am most grateful to Dr. P. Winternitz for many interesting discussions on this topic.

APPENDIX A

The Cross-Basis ME's

It is convenient to use for these a representation other than (4). We take

$$D^{jt}_{m\mu}(r) = (f_m, T^j_r f^t_\mu),$$

where

$$f_m = u^{-1}\colon \psi_m, \quad f^t_\mu = v^{-1}\colon \phi^t_\mu,$$

$$f_m(x) = (1+x^2)^j \left\{\frac{1-ix}{1+ix}\right\}^m (2\pi)^{-\frac{1}{2}},$$

$$f^t_\mu(x) = |1-x^2|^j \left|\frac{1+x}{1-x}\right|^{i\mu} (-\operatorname{sgn}\chi)^{\omega r} \times (2\pi)^{-\frac{1}{2}}\theta(t - t|x|). \tag{A1}$$

We can then write (for both signs of ξ)

$$D^{j+}_{m\mu}(\xi) = \frac{e^{-j\xi}}{\pi}\int_{-e^{-\xi}}^{e^{-\xi}} (1+x^2)^{-j-1}(1-e^{2\xi}x^2)^j \times \left\{\frac{1+ix}{1-ix}\right\}^m \left\{\frac{1+e^{\xi}x}{1-e^{\xi}x}\right\}^{i\mu} dx, \tag{A2}$$

and for $\xi > 0$ this is easily seen to be proportional to the analytic continuation of functions already known:

$$D^{j+}_{m\mu}(\xi) = e^{-\frac{1}{2}i\pi(j+1)} D^{j++}_{im,\mu}(\xi - \tfrac{1}{2}i\pi). \tag{A3}$$

Since (A2) defines a function of ξ which is analytic at $\xi = 0$, the same holds for negative ξ, with the proviso that the right-hand side of (A3) is continued from that functional form of $D^{j++}_{\mu'\mu}(\xi)$ valid for $\xi > 0$.

To find $D^{j-}(\xi)$, we use the relation $\epsilon\xi = (-\xi)\epsilon$, together with

$$T^j_\epsilon\colon \psi_m = e^{im\pi}\psi_m,$$
$$T^j_\epsilon\colon \phi^+_\mu = (-1)^\omega \phi^-_{-\mu}, \tag{A4}$$

and obtain

$$D^{j-}_{m\mu}(\xi) = e^{-im\pi} D^{j+}_{-m\mu}(-\xi), \tag{A5}$$

which result can also be obtained by direct calculation after the manner of (A2). By manipulating the integrals, we find the identities

$$D^{jt}_{m\mu}(\xi) = (-1)^{\omega r} D^{jt}_{-m-\mu}(\xi) = (-1)^{\omega r}\overline{D^{-j-1t}_{m\mu}}(\xi). \tag{A6}$$

Thus the overlap functions $D^{jt}_{m\mu}(0)$ are essentially just the matrix elements of a boost through $i\pi/2$: Compare this with the results of Ref. 7 for those of $SL(2, C)$. Because the explicit form of the D-functions contains half-angles, it is advantageous to make the substitution

$$\tanh \xi/2 = \tan \phi/2, \tag{A7}$$

so that

$$-\pi/2 < \phi < \pi/2;$$

then we can write explicitly

$$D^{j+}_{m\mu}(\xi) = \frac{\Gamma(j+i\mu+1)\Gamma(j-i\mu+1)}{2\pi\Gamma(2j+2)} e^{i\pi/2(j+1)} \times e^{i(\phi-\pi/2)(m-i\mu)}(2e^{-i\phi}\cos\phi)^{j+1} \times F(j+i\mu+1, j-m+1; 2j+2; 2e^{-i\phi}\cos\phi). \tag{A8}$$

This is the final expression. Notice that for $\phi > 0$ (that is, for $\xi > 0$) we take the principal branch of the hypergeometric function, but for $\phi < 0$ it passes through the cut onto the second sheet. For $m = 0$ it

reduces[8] to the well-known functions

$$D^{j\pm}_{0\mu}(\xi) = \frac{\Gamma(j + i\mu + 1)\Gamma(j - i\mu + 1)}{\Gamma(j + 1)} (2\pi)^{-\frac{1}{2}} \times (\cosh \xi)^{-\frac{1}{2}} P^{-j-\frac{1}{2}}_{i\mu-\frac{1}{2}}(\pm \tanh \xi), \quad \text{(A9)}$$

which play the role of the spherical harmonics[4] in the hyperbolic parametrization.

The asymptotic behavior of these functions depends upon the sign of ξ; we can see this most clearly by expressing the second-sheet values of F in (A8) in terms of its first-sheet ones, when we obtain

$$D^{j+}_{m\mu}(\xi) = c_1 \tilde{D}^{j+}_{m\mu}(-\xi) + c_2 \tilde{D}^{-j-1+}_{m\mu}(-\xi),$$

where the $\tilde{D}$ are given by (A8) evaluated on the principal branch. This agrees with the behavior of (A9), which we could derive by a quadratic transformation of the hypergeometric equation, and shows that for $t\xi > 0$ the behavior is simple (that is, second-kind in j), whereas for $\xi t < 0$ it is not.

We can make use of the equivalence of the representations j and $-j - 1$ to obtain an interesting identity. Recall our discussion in I of the intertwining operator $A: D_j \to D_{-j-1}$ and its coefficients

$$\begin{aligned} \alpha^l_\mu &= \Gamma(j + i\mu + 1)\Gamma(j - i\mu + 1) \\ &\quad \times \cos \pi(i\mu - \omega/2)/\pi, \\ \alpha'^l_\mu &= -\Gamma(j + i\mu + 1)\Gamma(j - i\mu + 1) \\ &\quad \times \cos \pi(j - \omega/2)/\pi; \end{aligned} \quad \text{(A10)}$$

by a short calculation we find that with the same normalization we have $A: \psi^j_m \to a_m \psi^{-j-1}_m$, where

$$a_m = \Gamma(j + m + 1)\Gamma(j - m + 1) \times \sin \pi(m - j) e^{-i\pi\omega/2}/\pi; \quad \text{(A11)}$$

thus the relation

$$(\psi^j_m, S^j_\xi : \phi^{jt}_\mu) = (A\psi^j_m, S^{-j-1}_\xi : A\phi^{jt}_\mu)$$

gives us the two equations

$$\begin{aligned} \alpha'_\mu D^{-j-1-}_{m\mu}(\xi) &= a_m D^{j+}_{m\mu}(\xi) - \alpha_\mu D^{-j-1+}_{m\mu}(\xi), \\ (-1)^\omega \alpha'_{-\mu} D^{-j-1+}_{m\mu}(\xi) &= a_m D^{j-}_{m\mu}(\xi) - \alpha_{-\mu} D^{-j-1-}_{m\mu}(\xi), \end{aligned} \quad \text{(A12)}$$

which are valid on the principal series. It is not difficult to show that with the specific form of the intertwining coefficients one of these implies the other; by using (A8), taking care to evaluate the hypergeometric functions on the correct sheet,[8] we have succeeded in verifying them analytically.

APPENDIX B

Calculation of $D^\chi(\theta')$

With the parametrization $\epsilon^\rho \alpha \xi \beta \theta' = k \epsilon^{\rho'} \alpha' \xi' \beta'$ we obtain

$$\tanh \xi'/2 = \frac{\tanh \xi/2 - \tan \theta/2}{1 + \tanh \xi/2 \tan \theta/2}, \quad \text{(B1)}$$

$$|\lambda^2| = \cos \theta + \sin \theta \sinh \xi \quad \text{(B2)}$$

when $\rho = \rho' = 0$; so that upon setting $\tanh \xi/2 = \tan \phi/2$ and regarding the basis functions as functions of ϕ we find

$$D^{\chi++}_{l\mu t, j\mu t'}(\theta') = \tfrac{1}{2} \int_{\theta-\pi/2}^{\pi/2} \overline{D^{lt}_{j_0\mu}(\phi)} \left\{ \frac{\cos(\phi - \theta)}{\cos \phi} \right\}^{\sigma-1} \times D^{jt'}_{j_0\mu}(\phi - \theta) \sec^2 \phi \, d\phi. \quad \text{(B3)}$$

To evaluate this, it is not convenient to use the form (A8) of the basis functions; instead we make use of[8] HTF 2.9(34) and obtain

$$\begin{aligned} D^{j+}_{m\mu}(\phi) &= \frac{\Gamma(j + i\mu + 1)\Gamma(-m - i\mu)}{2\pi\Gamma(j - m + 1)} e^{-i\pi/2(j+1)} \\ &\quad \times e^{i(\phi-\pi/2)(m+i\mu)} (2e^{i\phi} \cos \phi)^{j+1} \\ &\quad \times F(j + i\mu + 1, j + m + 1; \\ &\qquad m + i\mu + 1; -e^{2i\phi}) \\ &\quad + (m, \mu \to -m, -\mu), \end{aligned} \quad \text{(B4)}$$

and notice that we can take the principal branch of this function for all ξ. The hypergeometric series converges conditionally for all real $\phi \neq (n + 1)\pi$, and if we let $\operatorname{Re} j \to -\frac{1}{2} - \epsilon$ it does so absolutely. Using (A6) for the complex conjugation and letting $\operatorname{Re} l \to -\frac{1}{2} + \epsilon$, we can integrate term by term to obtain at last

$$\begin{aligned} &D^{\chi++}_{l\mu+, j\mu+}(\theta') \\ &= \frac{\Gamma(j + \sigma + 1)\Gamma(-l - \sigma)}{\Gamma(j - l + 1)} \\ &\quad \times \sum \frac{\Gamma(j+i\mu+1)\Gamma(m+i\mu)\Gamma(-l+i\mu')\Gamma(-m'-i\mu')}{2\pi^2\Gamma(j-m+1)\Gamma(-l+m')} \\ &\quad \times e^{i\theta(j-l+\sigma+1+m'+i\mu')} e^{-i\pi(2j-l+\sigma+2m_1+2i\mu_1)} \\ &\quad \times (2 \sin \theta)^{j-l} \sum_t \frac{(i\mu' - l)_t (m' - l)_t}{(m' + i\mu' + 1)_t t!} e^{2it\theta} \\ &\quad \times F(j - l + t + m_1 + i\mu_1, j + \sigma + 1; \\ &\qquad j - l + 1; 1 - e^{2i\theta}) \\ &\quad \times {}_4F_3(j + i\mu + 1, j + m + 1, -m' - i\mu' - t, \\ &\qquad -t; m + i\mu + 1, l - m' - t + 1, \\ &\qquad l - \mu' - t + 1, e^{-2i\theta}), \end{aligned} \quad \text{(B5)}$$

where

$$2m_1 = m + m',$$
$$2\mu_1 = \mu + \mu',$$
$$|m\mu| = |m'\mu'| = |j_0\mu|,$$

and the first summation is over all four possible combinations of signs of m, m', μ, μ' when those of the products $(m\mu)$, $(m'\mu')$ are held fixed.

Now let us consider D^{-+} $(0 < \theta' < \pi/2)$. We find

$$\alpha\xi\beta\theta' = k\epsilon(-\alpha - \pi)\xi'\beta,$$
$$\lambda^2 = \cos(\theta - \phi - \pi)/\cos\phi,$$

where

$$\tan\tfrac{1}{2}\phi' = \cot(\phi - \theta)/2, \qquad \text{(B6)}$$

with the notation used previously. The equation for ϕ' has the solution

$$\phi' = \theta - \phi - \pi,$$

and we find

$$D^{\chi-+}_{l\mu t, j\mu t'}(\theta') = \frac{e^{-i\pi j_0}}{2}\int_{-\pi/2}^{\theta-\pi/2} \overline{D^{lt}_{-j_0\mu}(\phi)}\left\{\frac{\cos(\theta - \phi - \pi)}{\cos\phi}\right\}^{\sigma-1} \times D^{jt'}_{j_0\mu}(\theta - \phi - \pi)\frac{d\phi}{\cos^2\phi}, \qquad \text{(B7)}$$

where both basis functions have their arguments in the meaningful interval $(-\pi/2, \pi/2)$. By using (A5) we find that formally this becomes

$$D^{\chi-+}_{l\mu t, j\mu t'}(\theta') = e^{2\pi i j_0} D^{\chi++}_{l\mu -t, j\mu t'}(\theta' \to \pi - \theta'), \qquad \text{(B8)}$$

where the right-hand side is to be interpreted as the analytic continuation to the region $\frac{1}{2}\pi < \theta < \pi$ of the functional form of the matrix element $D(\theta)$ appropriate for $0 < \theta < \frac{1}{2}\pi$. By using this, together with the identities (25) (which are derived by similar manipulations), we can obtain explicit expressions for all the matrix elements of this coset class.

[1] N. W. Macfadyen, Carnegie-Mellon University preprint (to be published). Hereafter referred to as I. See also *Proceedings of the Boulder Symposium on the De Sitter and Conformal Groups, University of Colorado, Summer* 1970.
[2] I. M. Gel'fand, M. I. Graev, and N. Ya. Vilenkin, *Generalized Functions* (Academic, New York, 1966), Vol. 5.
[3] W. Rühl, *The Lorentz Groups and Harmonic Analysis* (Benjamin, New York, 1969).
[4] N. Ya. Vilenkin and Ya. A. Smorodinskii, Zh. Eksp. Teor. Fiz. **46**, 1793 (1964) [Sov. Phys.—JETP **19**, 1209 (1964)].
[5] V. S. Popov, Zh. Eksp. Teor. Fiz. **37**, 1116 (1959) [Sov. Phys.—JETP **10**, 794 (1960)].
[6] G. I. Kuznetsov, Ya. A. Smorodinskii, M. A. Liberman, and A. A. Makarov, Yad. Fiz. **10**, 644 (1969) [Sov. J. Nucl. Phys. **10**, 370 (1970)].
[7] R. Delbourgo, K. Koller, and P. Mahanta, Nuovo Cimento **52A**, 1254 (1967).
[8] A. Erdélyi *et al.*, *Higher Transcendental Functions* (McGraw-Hill, New York, 1953), Vol. 1.

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 12, NUMBER 3 MARCH 1971

The Einstein Tensor and Its Generalizations*

DAVID LOVELOCK
Department of Applied Mathematics, University of Waterloo, Waterloo, Ontario, Canada

(Received 27 August 1970)

The Einstein tensor G^{ij} is symmetric, divergence free, and a concomitant of the metric tensor g_{ab} together with its first two derivatives. In this paper all tensors of valency two with these properties are displayed explicitly. The number of independent tensors of this type depends crucially on the dimension of the space, and, in the four dimensional case, the only tensors with these properties are the metric and the Einstein tensors.

1. INTRODUCTION

In most introductions to the general theory of relativity, the motivation which gives rise to the Einstein field equations *in vacuo* usually involves solving the following problem: to seek all tensors A^{ij} with the properties:

(a) A^{ij} is symmetric, i.e.,[1]

$$A^{ij} = A^{ji};$$

(b) A^{ij} is a concomitant of the metric tensor g_{ab} and its first two derivatives, i.e.,[2]

$$A^{ij} = A^{ij}(g_{ab}; g_{ab,c}; g_{ab,cd});$$

(c) A^{ij} is divergence free, i.e.,[3]

$$A^{ij}{}_{|j} = 0;$$

(d) A^{ij} is linear in the second derivatives of g_{ab}.

The field equations *in vacuo* are then assumed to take the form

$$A^{ij} = 0.$$

where

$$2m_1 = m + m',$$
$$2\mu_1 = \mu + \mu',$$
$$|m\mu| = |m'\mu'| = |j_0\mu|,$$

and the first summation is over all four possible combinations of signs of m, m', μ, μ' when those of the products $(m\mu)$, $(m'\mu')$ are held fixed.

Now let us consider D^{-+} $(0 < \theta' < \pi/2)$. We find

$$\alpha\xi\beta\theta' = k\epsilon(-\alpha - \pi)\xi'\beta,$$
$$\lambda^2 = \cos(\theta - \phi - \pi)/\cos\phi,$$

where

$$\tan\tfrac{1}{2}\phi' = \cot(\phi - \theta)/2, \qquad \text{(B6)}$$

with the notation used previously. The equation for ϕ' has the solution

$$\phi' = \theta - \phi - \pi,$$

and we find

$$D^{\chi-+}_{l\mu t, j\mu t'}(\theta') = \frac{e^{-i\pi j_0}}{2}\int_{-\pi/2}^{\theta-\pi/2} \overline{D^{lt}_{-j_0\mu}(\phi)}\left\{\frac{\cos(\theta - \phi - \pi)}{\cos\phi}\right\}^{\sigma-1} \times D^{jt'}_{j_0\mu}(\theta - \phi - \pi)\frac{d\phi}{\cos^2\phi}, \qquad \text{(B7)}$$

where both basis functions have their arguments in the meaningful interval $(-\pi/2, \pi/2)$. By using (A5) we find that formally this becomes

$$D^{\chi-+}_{l\mu t, j\mu t'}(\theta') = e^{2\pi i j_0} D^{\chi++}_{l\mu -t, j\mu t'}(\theta' \to \pi - \theta'), \qquad \text{(B8)}$$

where the right-hand side is to be interpreted as the analytic continuation to the region $\frac{1}{2}\pi < \theta < \pi$ of the functional form of the matrix element $D(\theta)$ appropriate for $0 < \theta < \frac{1}{2}\pi$. By using this, together with the identities (25) (which are derived by similar manipulations), we can obtain explicit expressions for all the matrix elements of this coset class.

[1] N. W. Macfadyen, Carnegie-Mellon University preprint (to be published). Hereafter referred to as I. See also *Proceedings of the Boulder Symposium on the De Sitter and Conformal Groups, University of Colorado, Summer* 1970.
[2] I. M. Gel'fand, M. I. Graev, and N. Ya. Vilenkin, *Generalized Functions* (Academic, New York, 1966), Vol. 5.
[3] W. Rühl, *The Lorentz Groups and Harmonic Analysis* (Benjamin, New York, 1969).
[4] N. Ya. Vilenkin and Ya. A. Smorodinskii, Zh. Eksp. Teor. Fiz. **46**, 1793 (1964) [Sov. Phys.—JETP **19**, 1209 (1964)].
[5] V. S. Popov, Zh. Eksp. Teor. Fiz. **37**, 1116 (1959) [Sov. Phys.—JETP **10**, 794 (1960)].
[6] G. I. Kuznetsov, Ya. A. Smorodinskii, M. A. Liberman, and A. A. Makarov, Yad. Fiz. **10**, 644 (1969) [Sov. J. Nucl. Phys. **10**, 370 (1970)].
[7] R. Delbourgo, K. Koller, and P. Mahanta, Nuovo Cimento **52A**, 1254 (1967).
[8] A. Erdélyi *et al.*, *Higher Transcendental Functions* (McGraw-Hill, New York, 1953), Vol. 1.

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 12, NUMBER 3 MARCH 1971

The Einstein Tensor and Its Generalizations*

DAVID LOVELOCK
Department of Applied Mathematics, University of Waterloo, Waterloo, Ontario, Canada

(Received 27 August 1970)

The Einstein tensor G^{ij} is symmetric, divergence free, and a concomitant of the metric tensor g_{ab} together with its first two derivatives. In this paper all tensors of valency two with these properties are displayed explicitly. The number of independent tensors of this type depends crucially on the dimension of the space, and, in the four dimensional case, the only tensors with these properties are the metric and the Einstein tensors.

1. INTRODUCTION

In most introductions to the general theory of relativity, the motivation which gives rise to the Einstein field equations *in vacuo* usually involves solving the following problem: to seek all tensors A^{ij} with the properties:

(a) A^{ij} is symmetric, i.e.,[1]

$$A^{ij} = A^{ji};$$

(b) A^{ij} is a concomitant of the metric tensor g_{ab} and its first two derivatives, i.e.,[2]

$$A^{ij} = A^{ij}(g_{ab}; g_{ab,c}; g_{ab,cd});$$

(c) A^{ij} is divergence free, i.e.,[3]

$$A^{ij}{}_{|j} = 0;$$

(d) A^{ij} is linear in the second derivatives of g_{ab}.

The field equations *in vacuo* are then assumed to take the form

$$A^{ij} = 0.$$

Of these properties, (d) is usually regarded as crucial, and it is certainly essential when, as is customary, appeal is made to the work of either Cartan[4] or Weyl and Vermeil[5] in order to conclude that

$$A^{ij} = aG^{ij} + bg^{ij},$$

where a and b are constants and G^{ij} is the Einstein tensor.[6] The resulting field equations would then be the Einstein equations with the cosmological term. However, none of these remarks depend on the dimension n of the underlying Riemannian space.

Recently, the problem of finding all tensors A^{ij} which satisfy (a), (b), and (c) [without insisting on (d)] has been investigated.[7] Although the general structure of A^{ij} was found, it was given in terms of an iterative procedure, and no explicit formula for A^{ij} was available. We briefly summarize these results.

If we define the positive integer m by

$$m = \tfrac{1}{2}n, \quad \text{if } n \text{ is even},$$
$$= \tfrac{1}{2}(n+1), \quad \text{if } n \text{ is odd},$$

then[8] the most general tensor A^{ij} satisfying (a), (b), and (c) is given by

$$A^{ij} = \sum_{p=1}^{m-1} c_p \theta^{ij;i_1i_2\cdots i_{4p-1}i_{4p}} \prod_{t=1}^{p} R_{i_{4t-1}i_{4t-3}i_{4t-2}i_{4t}} + ag^{ij}, \tag{1.1}$$

where c_p and a are constants and $\theta^{ij;i_1\cdots i_{4p}}$, $p = 1, \cdots, m-1$, are tensors with the following properties:

(i) They are concomitants of g_{ab}, i.e.,

$$\theta^{ij;i_1\cdots i_{4p}} = \theta^{ij;i_1\cdots i_{4p}}(g_{ab});$$

(ii) they are symmetric in ij and in $i_{2t-1}i_{2t}$ for $t = 1, 2, \cdots, 2p$;

(iii) they are symmetric under interchange of the pair (i,j) with the pair (i_{2t-1}, i_{2t}) for all $t = 1, 2, \cdots, 2p$;

(iv) they satisfy the cyclic identity involving any three of the four indices $(i,j)(i_{2t-1}, i_{2t})$ for $t = 1, 2, \cdots, 2p$, e.g.,

$$\theta^{ij;i_1i_2\cdots i_{4p}} + \theta^{i_1i;ji_2\cdots i_{4p}} + \theta^{ji_1;ii_2\cdots i_{4p}} = 0.$$

[We remark that (ii), (iii), and (iv) are not independent—in fact, (ii) and (iv) imply (iii).]

A complicated and unwieldy iterative formula was given[9] for calculating $\theta^{ij;i_1\cdots i_{4p}}$ for $p = 1, \cdots, m-1$, which we shall not repeat here. However, for fixed p this procedure establishes that (i)–(iv) determine $\theta^{ij,i_1\cdots i_{4p}}$ uniquely up to a constant. This uniqueness ensures that, if we can supply a nonzero tensor which satisfies (i)–(iv), then (up to a constant which can be absorbed in c_p) we may use it in (1.1) to calculate A^{ij} explicitly. The purpose of this paper is to carry through this program. The resulting form for A^{ij} is displayed explicitly in a manner which very clearly demonstrates the role played by the dimension.

2. DIVERGENCE-FREE TENSORIAL CONCOMITANTS

For any positive integer p, $1 \leq p \leq m-1$, we define the tensor

$$\psi^{ij;i_1i_2\cdots i_{4p-1}i_{4p}} = (\delta^{ih_1h_2\cdots h_{2p}}_{kj_1j_2\cdots j_{2p}} g^{kj} + \delta^{jh_1h_2\cdots h_{2p}}_{kj_1j_2\cdots j_{2p}} g^{ki}) g^{j_1k_1} g^{j_2k_2} \cdots g^{j_{2p}k_{2p}} \times D^{i_1i_2i_3i_4}_{h_1h_2k_1k_2} \cdots D^{i_{4p-3}i_{4p-2}i_{4p-1}i_{4p}}_{h_{2p-1}h_{2p}k_{2p-1}k_{2p}}, \tag{2.1}$$

where $\delta^{i_1i_2\cdots i_N}_{j_1j_2\cdots j_N}$ is the generalized Kronecker delta defined by

$$\delta^{i_1i_2\cdots i_N}_{j_1j_2\cdots j_N} = \det \begin{vmatrix} \delta^{i_1}_{j_1} & \cdots & \delta^{i_1}_{j_N} \\ \cdot & & \cdot \\ \cdot & & \cdot \\ \cdot & & \cdot \\ \delta^{i_N}_{j_1} & \cdots & \delta^{i_N}_{j_N} \end{vmatrix}$$

and

$$D^{ijkl}_{abcd} = \tfrac{1}{2}(\delta^i_a\delta^j_d + \delta^i_d\delta^j_a)(\delta^k_b\delta^l_c + \delta^k_c\delta^l_b).$$

We will now demonstrate that $\psi^{ij;i_1\cdots i_{4p}}$ has the properties (i)–(iv) which we require, and is thus $\theta^{ij;i_1\cdots i_{4p}}$ up to a constant. We shall first show that $\psi^{ij;i_1\cdots i_{4p}}$ is nonzero by proving that the same is true for $\psi^{ij;i_1\cdots i_{4p}} g_{ij}g_{i_1i_2}\cdots g_{i_{4p-1}i_{4p}}$. Clearly, we have

$$\psi^{ij;i_1\cdots i_{4p}} g_{ij}g_{i_1i_2}\cdots g_{i_{4p-1}i_{4p}} = (-1)^p 2^{2p+1} \delta^{kh_1\cdots h_{2p}}_{kh_1\cdots h_{2p}}.$$

If $n \geq 2p+1$ (which is always satisfied for $1 \leq p \leq m-1$), the right-hand side of the equation reduces to

$$(-1)^p 2^{p+1} n!/(n-2p-1)!,$$

which is nonzero.

Quite clearly conditions (i) and (ii) are satisfied by $\psi^{ij;i_1\cdots i_{4p}}$. We now turn to the remaining conditions (iii) and (iv) [of which (iii) is a consequence of (ii) and (iv)]. In view of the fact that $\psi^{ij;i_1\cdots i_{4p}}$ is symmetric under interchange of the indices $i_1i_2i_3i_4$ with the indices $i_{4h-3}i_{4h-2}i_{4h-1}i_{4h}$ for any h, $2 \leq h \leq m-1$, it suffices to prove (iv) for the pairs ij and i_1i_2. (We need not consider the pairs ij and i_3i_4 in view of the obvious symmetry of $\psi^{ij;i_1\cdots i_{4p}}$ under interchange of i_1i_2 with i_3i_4.) The fact that (iv) is satisfied is now trivially established. Consequently, in view of the uniqueness, we have

$$\theta^{ij;i_1\cdots i_{4p}} = b_p \psi^{ij;i_1\cdots i_{4p}} \tag{2.2}$$

for $1 \leq p \leq m-1$, where the b_p are constants.

We now wish to substitute (2.2) and (2.1) in (1.1), which will clearly involve expressions like

$$D^{i_{4t-3}i_{4t-2}i_{4t-1}i_{4t}}_{h_{2t-1}h_{2t}k_{2t-1}k_{2t}} R_{i_{4t-1}i_{4t-3}i_{4t-2}i_{4t}},$$

A trivial calculation shows that

$$D^{i_{4t}-3i_{4t}-2i_{4t}-1i_{4t}}_{h_{2t-1}h_{2t}k_{2t-1}k_{2t}} R_{i_{4t-1}i_{4t-3}i_{4t-2}i_{4t}} = R_{h_{2t-1}h_{2t}k_{2t-1}k_{2t}} + R_{h_{2t-1}k_{2t-1}h_{2t}k_{2t}}.$$

If we interchange h_{2t-1} and h_{2t} and subtract, we find

$$(D^{i_{4t}-3i_{4t}-2i_{4t}-1i_{4t}}_{h_{2t-1}h_{2t}k_{2t-1}k_{2t}} - D^{i_{4t}-3i_{4t}-2i_{4t}-1i_{4t}}_{h_{2t}h_{2t-1}k_{2t-1}k_{2t}}) R_{i_{4t-1}i_{4t-3}i_{4t-2}i_{4t}} = 3R_{h_{2t-1}h_{2t}k_{2t-1}k_{2t}} \quad (2.3)$$

by virtue of the antisymmetry and cyclic properties of the curvature tensor. By noting that (2.1) can be expressed in the form

$$\psi^{ij;i_1\cdots i_{4p}} = (2^{-p})(\delta^{ih_1\cdots h_{2p}}_{kj_1\cdots j_{2p}} g^{kj} + \delta^{jh_1\cdots h_{2p}}_{kj_1\cdots j_{2p}} g^{ki}) \times g^{j_1k_1}\cdots g^{j_{2p}k_{2p}} (D^{i_1i_2i_3i_4}_{h_1h_2k_1k_2} - D^{i_1i_2i_3i_4}_{h_2h_1k_1k_2})\cdots (D^{i_{4p}-3i_{4p}-2i_{4p}-1i_{4p}}_{h_{2p-1}h_{2p}k_{2p-1}k_{2p}} - D^{i_{4p}-3i_{4p}-2i_{4p}-1i_{4p}}_{h_{2p}h_{2p-1}k_{2p-1}k_{2p}}), \quad (2.4)$$

we see that (1.1), (2.2), (2.3), and (2.4) imply that

$$A^{ij} = \sum_{p=1}^{m-1} a_p(\delta^{ih_1\cdots h_{2p}}_{kj_1\cdots j_{2p}} g^{kj} + \delta^{jh_1\cdots h_{2p}}_{kj_1\cdots j_{2p}} g^{ki}) \times \prod_{t=1}^{p} R_{h_{2t-1}h_{2t}}{}^{j_{2t-1}j_{2t}} + ag^{ij},$$

where a and a_p are constants. It is now easy to prove the following.

Theorem 1: The only symmetric tensor $A^{ij} = A^{ij}(g_{rs}; g_{rs,t}; g_{rs,tu})$ for which

$$A^{ij}{}_{|j} \equiv 0$$

is

$$A^i_j = \sum_{p=1}^{m-1} a_p \delta^{ih_1\cdots h_{2p}}_{jj_1\cdots j_{2p}} R_{h_1h_2}{}^{j_1j_2} R_{h_3h_4}{}^{j_3j_4}\cdots R_{h_{2p-1}h_{2p}}{}^{j_{2p-1}j_{2p}} + a\delta^i_j, \quad (2.5)$$

where a and a_p are arbitrary constants.

For $n = 4$, the case of interest in general relativity, (2.5) reduces to

$$A^i_j = a\delta^i_j + b\delta^{ii_1i_2}_{jj_1j_2} R_{i_1i_2}{}^{j_1j_2} = a\delta^i_j - 4bG^i_j.$$

We have thus recovered[10] the following:

Corollary: If $n = 4$ the only symmetric tensor $A^{ij} = A^{ij}(g_{rs}; g_{rs,t}; g_{rs,tu})$ for which

$$A^{ij}{}_{|j} \equiv 0$$

is

$$A^{ij} = aG^{ij} + bg^{ij}.$$

Consequently, if $n = 4$, (a), (b), and (c) imply (d) so that in general relativity this apparently crucial assumption is not required.

By virtue of the fact that[11] if $n < N$, then

$$\delta^{i_1i_2\cdots i_N}_{j_1j_2\cdots j_N} = 0 \quad \text{identically;} \quad (2.6)$$

we can express (2.5) as an infinite series, viz.,

$$A^i_j = \sum_{p=1}^{\infty} a_p \delta^{ii_1\cdots i_{2p}}_{jj_1\cdots j_{2p}} R_{i_1i_2}{}^{j_1j_2}\cdots R_{i_{2p-1}i_{2p}}{}^{j_{2p-1}j_{2p}} + a\delta^i_j, \quad (2.7)$$

which, in fact, has only a finite number of terms depending on the dimension n of the space.

3. THE ASSOCIATED LAGRANGE DENSITY

A natural question now arises: Is[12] $g^{\frac{1}{2}}A^{ij}$ the Euler–Lagrange expression of a suitably chosen Lagrange density? Elsewhere[13] we have shown that this question can be answered in the affirmative. However, the argument used was involved, this complication being directly due to the fact that an explicit form for A^{ij} was not available. From (2.5) it is a simple matter to prove the following.

Theorem 2: If A^{ij} is given by (2.5), then $g^{\frac{1}{2}}A^{ij}$ is the Euler–Lagrange expression corresponding to[14]

$$L = g^{\frac{1}{2}} \sum_{p=1}^{m-1} 2a_p \delta^{h_1\cdots h_{2p}}_{j_1\cdots j_{2p}} R_{h_1h_2}{}^{j_1j_2}\cdots R_{h_{2p-1}h_{2p}}{}^{j_{2p-1}j_{2p}} + 2ag^{\frac{1}{2}}. \quad (3.1)$$

Proof: It has been shown by Rund[15] that the Euler–Lagrange expression E^{ij}, corresponding to a Lagrange density

$$L = L(g_{ij}; g_{ij,k}; g_{ij,kh}),$$

viz.,

$$E^{ij} = \frac{\partial L}{\partial g_{ij}} - \frac{\partial}{\partial x^h}\left(\frac{\partial L}{\partial g_{ij,h}}\right) + \frac{\partial^2}{\partial x^h \partial x^k}\left(\frac{\partial L}{\partial g_{ij,hk}}\right), \quad (3.2)$$

can be expressed in the tensorial form

$$E^{ij} = \Lambda^{ij,hk}{}_{|hk} + \tfrac{1}{2}g^{ij}L - \tfrac{2}{3}R_k{}^j{}_{lh}\Lambda^{hk,il}, \quad (3.3)$$

where

$$\Lambda^{ij,hk} = \frac{\partial L}{\partial g_{ij,hk}}. \quad (3.4)$$

Consequently, by means of (3.3), a knowledge of $\Lambda^{ij,hk}$ alone will enable us to calculate E^{ij} [whereas, if we were to use (3.2), not only (3.4) but also $\partial L/\partial g_{ij}$ and $\partial L/\partial g_{ij,h}$ are required].

From (3.4) and (3.1) we find

$$\Lambda^{ab,cd} = g^{\frac{1}{2}} \sum_{p=1}^{m-1} 2pa_p \delta^{i_1 \cdots i_{2p}}_{j_1 \cdots j_{2p}} R_{i_3 i_4}{}^{j_3 j_4} \cdots R_{i_{2p-1} i_{2p}}{}^{j_{2p-1} j_{2p}} g^{j_2 t} g^{j_1 u} D^{abcd}_{i_1 i_2 ut}, \quad (3.5)$$

where, as before,

$$D^{abcd}_{ijut} = \tfrac{1}{2}(\delta^a_i \delta^b_t + \delta^a_t \delta^b_i)(\delta^c_j \delta^d_u + \delta^c_u \delta^d_j).$$

Covariant differentiation of (3.5) with respect to x^c yields

$$\Lambda^{ab,cd}{}_{|c} = g^{\frac{1}{2}} \sum_{p=1}^{m-1} 2p(p-1) a_p \delta^{i_1 \cdots i_{2p}}_{j_1 \cdots j_{2p}} R_{i_3 i_4}{}^{j_3 j_4}{}_{|c} \times R_{i_5 i_6}{}^{j_5 j_6} \cdots R_{i_{2p-1} i_{2p}}{}^{j_{2p-1} j_{2p}} g^{j_1 u} g^{j_2 t} D^{abcd}_{i_1 i_2 ut}.$$

A simple argument, involving the use of the Bianchi identities, shows that[16] for (3.1)

$$\Lambda^{ab,cd}{}_{|c} = 0. \quad (3.6)$$

If we introduce the notation

$$B^{i_1 i_2}_{j_1 j_2} = g^{\frac{1}{2}} \delta^{i_1 i_2 \cdots i_{2p}}_{j_1 j_2 \cdots j_{2p}} R_{i_3 i_4}{}^{j_3 j_4} \cdots R_{i_{2p-1} i_{2p}}{}^{j_{2p-1} j_{2p}},$$

then, from (3.5),

$$R_{bkda} \Lambda^{ab,cd} = \sum_{p=1}^{m-1} 2pa_p B^{ij}_{rs} g^{ru} g^{st} D^{abcd}_{ijut} R_{bkda}. \quad (3.7)$$

It is easily shown that

$$B^{ij}_{rs} g^{ru} g^{st} D^{abcd}_{ijut} R_{bkda} = \tfrac{3}{2} B^{ic}_{rs} R_{ik}{}^{rs}, \quad (3.8)$$

so that substitution of (3.6), (3.7), and (3.8) in (3.3) gives

$$E^i_j = g^{\frac{1}{2}} \left(\sum_{p=1}^{m-1} a_p (\delta^i_j \delta^{h_1 \cdots h_{2p}}_{j_1 \cdots j_{2p}} R_{h_1 h_2}{}^{j_1 j_2} \cdots R_{h_{2p-1} h_{2p}}{}^{j_{2p-1} j_{2p}} - 2p \delta^{i h_2 \cdots h_{2p}}_{j_1 \cdots j_{2p}} R_{j h_2}{}^{j_1 j_2} \cdots R_{h_{2p-1} h_{2p}}{}^{j_{2p-1} j_{2p}}) + a \delta^i_j \right).$$

The right-hand side of the latter is clearly $g^{\frac{1}{2}} A^i_j$, which proves the theorem.

Finally, we remark that it is not true that (3.1) is the most general Lagrange density giving rise to (2.5). For example, if $n = 4$, the most general scalar density

$$L = L(g_{ij}; g_{ij,k}; g_{ij,kh}),$$

which has $g^{\frac{1}{2}} A^{ij}$ as its Euler–Lagrange expression, is[17]

$$g^{\frac{1}{2}} (\alpha + \beta \delta^{ij}_{ab} R_{ij}{}^{ab} + \gamma \delta^{ijkl}_{abcd} R_{ij}{}^{ab} R_{kl}{}^{cd}) + \mu R_{ijkl} {}^*R^{ijkl},$$

where α, β, γ, and μ are constants and ${}^*R^{ijkl}$ is the dual of R_{ijkl}.

* This work was supported in part by a research grant from the National Research Council of Canada.

[1] Unless otherwise specified, Latin indices run from 1 to n.

[2] A comma denotes partial differentiation.

[3] The summation convention is used throughout. The vertical bar denotes covariant differentiation.

[4] E. Cartan, J. Math. Pure Appl. **1**, 141 (1922).

[5] H. Weyl, *Space–Time–Matter* (Dover, New York, 1922), 4th ed., pp. 315ff; H. Vermeil, Nachr. Ges. Wiss. Göttingen, 334 (1917).

[6] If X^i is any contravariant vector field, then we define the Riemann curvature tensor $R_h{}^i{}_{jk}$, the Ricci tensor R_{hj}, the curvature scalar R, and the Einstein tensor G_{ij} by

$$X^i{}_{|jk} - X^i{}_{|kj} = R_h{}^i{}_{jk} X^h, \quad R_{kj} = R_h{}^i{}_{ji},$$
$$R = g^{hj} R_{hj}, \quad \text{and} \quad G_{ij} = R_{ij} - \tfrac{1}{2} g_{ij} R,$$

respectively.

[7] D. Lovelock, Aequationes Math. **4**, 127 (1970).

[8] Reference 7, Theorem 4.

[9] Reference 7, Theorem 3.

[10] Reference 7, Corollary 1.

[11] For various applications of (2.6), see D. Lovelock, Atti Accad. Nazl. Lincei **42**, 187 (1967); Proc. Cambridge Phil. Soc. **68**, 345 (1970).

[12] $g = \det g_{ij}$. Without loss of generality, we may assume $g > 0$.

[13] Reference 7, Theorem 5.

[14] This scalar has arisen elsewhere in an entirely different context. H. Rund, "Curvature Invariants Associated with Sets of n Fundamental Forms of Hypersurfaces of n-Dimensional Riemannian Manifolds" [to appear in Tensor (1971)]. D. Lovelock, "Intrinsic Expressions for Curvatures of Even Order of Hypersurfaces in a Euclidean Space" [to appear in Tensor (1971)].

[15] H. Rund, Abhandl. Math. Sem. Univ. Hamburg **29**, 243 (1966).

[16] The relationship of (3.6) to Lagrangians which satisfy the Euler–Lagrange equations identically has been investigated by R. Pavelle (private communication).

[17] D. Lovelock, Arch. Ratl. Mech. Anal. **33**, 54 (1969).

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 12, NUMBER 3 MARCH 1971

Diffraction Radiation by a Line Charge Moving past a Comb: A Model of Radiation Losses in an Electron Ring Accelerator*

R. D. HAZELTINE AND M. N. ROSENBLUTH
The Institute for Advanced Study, Princeton, New Jersey 08540
AND
A. M. SESSLER
Lawrence Radiation Laboratory, University of California, Berkeley, California 94720

(Received 7 July 1970)

A calculation is given of the radiated-energy loss from a charged rod which moves at constant speed past an infinite set of parallel semi-infinite conducting plates of infinitesimal thickness, with the rod taken parallel to and at a fixed distance from the plate edges. The problem is analyzed using the Wiener–Hopf technique, and the resulting formulas are evaluated analytically in the limits of high rod speed and low rod speed, and compared with numerical evaluation over the full range of speeds.

I. INTRODUCTION

An electron ring accelerator accelerates heavy ions by trapping the ions in the potential well associated with a compact ring of relativistic electrons, and then accelerating the electrons by means of externally applied fields.[1] It is clear that the highly charged electron ring will, while being accelerated, radiate strongly because of its motion past the conducting surfaces of the acceleration column. Considerable theoretical effort has been devoted to determining the extent of this radiation, more than a dozen different calculations having been reported.[2]

The crucial point is the dependence of the ring radiation, at ultrarelativistic speeds, upon ring speed. If, for example, the radiation were to increase with increasing speed, then the efficiency of an electron ring accelerator would decrease with increasing energy and there would result—in practice—an upper limit to the energy of the accelerator. Thus the very development of electron ring accelerators hinged upon demonstration that they would not be limited by radiation loss at high energies.

It is easy to estimate the radiation due to *acceleration* of the electron ring and to see that—at least in the relativistic limit—it is quite small. The radiation which is not small is the *diffraction radiation* due to the motion of the ring near conducting surfaces. Crudely speaking, one could say that image charges are being accelerated and hence there is radiation. It suffices to calculate the energy radiated by a ring moving at constant speed.

If the ring is approximated by a charge Q, then the net energy gain per unit length of the structure can be written in the form

$$\Delta U = AQ - BQ^2.$$

For a charge moving at constant speed, A is proportional to the externally applied fields in the structure and is, clearly, the energy gain for an infinitesimal charge. The term BQ^2 is, by superposition, independent of the external fields on the structure; thus it may be calculated for an unexcited structure. It is simply the radiated energy loss of a charge Q moving at constant speed through the structure. The considerable theoretical effort, mentioned above, has been devoted to determining B which is, clearly, a function only of charge speed and the geometry of the accelerating structure.

The simplest model which has been considered is that of a charge passing through a closed cylindrical cavity. The radiation loss into the cavity was found to increase with increasing γ, where $\gamma = [1 - (v^2/c^2)]^{-\frac{1}{2}}$ and v is the charge speed and c the speed of light.[3]

It was suggested by Kolpakov and Kotov that a reasonable approximation to a cavity with entrance and exit ports will omit the radiation for modes with wavelengths less than the port dimensions. The radiation loss is then found to be γ independent at large γ.[4]

A wave-diffraction model was employed by Lawson to study, more carefully, the short wavelength modes which were eliminated in the Kolpakov–Kotov approximation. Lawson found that they contributed energy loss which increased as $\gamma^{\frac{1}{2}}$ at large γ, and this result was obtained independently by Courant.[5,6]

There remained the possibility that the radiation loss to an infinite periodic array was quite different from the loss to a single cavity. Voskresenskii and Bolotovskii had derived an expression for the energy loss by a charged rod moving past a periodic array of semi-infinite planes,[7] which they subsequently employed to show that asymptotically the radiation varied

as $1/\gamma$ at large γ.[8] A γ-independent asymptotic dependence was obtained by Kuznetsov and Rubin.[9] Numerical evaluation of the Voskresenskii–Bolotovskii formula gave energy loss which fitted rather well—up to $\gamma \approx 300$—a $\gamma^{-\frac{1}{2}}$ dependence.[10]

Thus it seemed likely that there was not a practical limit to the energy of an electron ring accelerator—at least up until exceedingly high energies—and development programs pressed ahead in four different laboratories. There remained, however, the question of reconciling the numerical results with the asymptotic evaluations, and this task is accomplished in this paper.

Also, clearly, the radiation loss had to be evaluated for structures which approximate actual acceleration columns. Keil has studied, numerically, a periodic array of cylindrical cavities connected by beam pipes.[11] His analysis—in contrast with the work on the planar problem—must be cut off at short wavelengths. He finds energy loss which is γ independent, for large γ. The neglect of small wavelengths is supported by a numerical indication of convergences, and also by the results obtained in this paper. The difference between $\gamma^{-\frac{1}{2}}$ and γ^0 dependence, at large γ, is presumably a result of infinite transverse structure dimensions vs finite transverse dimensions. Energy balance arguments, presented in Appendix B, show that in a finite structure the energy loss can not decrease with increasing γ.

Still outstanding, at the present time, are results for periodic structure of finite length and for slightly imperfect structures. Efforts are, however, being put into these problems.[12] Rigorous analytic results for periodic, finite transverse dimensional structures would be most valuable, and hence worth the considerable effort they probably will demand.

Specifically, in this paper we compute the radiative energy loss from a charged rod which moves at constant speed past an infinite set of parallel semi-infinite conducting plates. The plates are uniformly spaced a distance $2\pi L$ apart, and the rod moves in the direction of their common normal at a distance x_0 below the plates' edges, as depicted in Fig. 1. We take the y direction as being perpendicular to the plane of the figure; note that all fields and currents may be assumed to be independent of y.

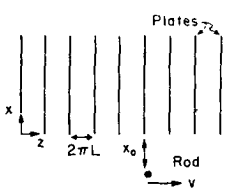

FIG. 1. Rod, plates, and coordinate system. The origin is at the edge of one of the plates.

Radiation problems with this boundary configuration were apparently first considered by Carlson and Heins.[13] However, the work of these authors, and some later studies by Heins,[14] did not consider our particular form of radiating source, and did not have occasion to compute energy losses. A problem identical with ours was analyzed by Voskresenskii and Botolovskii.[7,8] Despite this analysis, there are two reasons for reconsidering the problem here. First, as already stated, the work of Botolovskii and Voskresenskii is in conflict with the numerical evaluations.[10] Secondly, it would be desirable to have expressions for the energy loss, valid in the limiting regimes of low and ultrarelativistic rod speed, in which the dependence upon rod speed and geometrical parameters is transparent and from which numerical results readily may be obtained.

It is to this second task, i.e., the asymptotic evaluation of the energy loss, that our primary attention will be devoted. This is accomplished in Secs. 3 and 4. We first of all, however, derive in Sec. 2 the formal solution to the boundary value problem, both in order to correct an error in Ref. 7 and for the sake of completeness. Finally, the modifications required to treat a slightly different situation, in which the charged rod is replaced by a moving current, are briefly considered in Appendix A.

The main results of our analysis are the formal expression of Eqs. (34) and (36), the asymptotic formula of Eqs. (64)–(67), and the low-speed formula of Eq. (100) (with Fig. 5). Comparison between the asymptotic formulas, which have a dominant $\gamma^{-\frac{1}{2}}$ dependence, and direct numerical evaluation of Eqs. (34) and (36) is presented in Fig. 4. The results for a current-carrying rod are given in (A8) and (A9).

2. SOLUTION TO THE BOUNDARY VALUE PROBLEM

It follows from Maxwell's equations that the electric field $\mathcal{E}$ and current density $\mathbf{J}$ satisfy

$$\nabla \times (\nabla \times \mathcal{E}) = -4\pi \frac{\partial \mathbf{J}}{\partial t} - \frac{\partial^2 \mathcal{E}}{\partial t^2}. \tag{1}$$

(We use Gaussian units but set the light speed $c = 1$.) Here the left-hand side is

$$-\nabla^2 \mathcal{E} + \nabla(\nabla \cdot \mathcal{E}) = -\nabla^2 \mathcal{E} + \nabla(4\pi\rho).$$

But the charge density ρ satisfies $\partial\rho/\partial t + \nabla \cdot \mathbf{J} = 0$, so that, by differentiating Eq. (1) with respect to time, we may obtain an equation in which ρ does not appear:

$$\frac{\partial}{\partial t}\left[-\nabla^2 \mathcal{E} + \frac{\partial^2}{\partial t^2}\mathcal{E}\right] = 4\pi\left[\nabla(\nabla \cdot \mathbf{J}) - \frac{\partial^2 \mathbf{J}}{\partial t^2}\right]. \tag{2}$$

It suffices to solve the x component of this equation:

$$\frac{\partial}{\partial t}\left(\frac{\partial^2}{\partial t^2}-\frac{\partial^2}{\partial x^2}-\frac{\partial^2}{\partial z^2}\right)\mathcal{E}_x = 4\pi\left[\frac{\partial^2}{\partial x\partial z}J_z+\left(\frac{\partial^2}{\partial x^2}-\frac{\partial^2}{\partial t^2}\right)J_x\right]. \quad (3)$$

Here the unknowns are $\mathcal{E}_x$ and the induced surface current J_x, since J_z is given in terms of the motion of the charged rod. In fact, if the rod has speed v and charge per unit length q,

$$J_z = qv\delta(x+x_0)\delta(z-vt). \quad (4)$$

We have in addition the boundary condition that $\mathcal{E}_x$ vanish on the surface of each (infinitely conducting) plate,

$$\mathcal{E}_x|_{z=2\pi nL} = 0, \quad x>0, \quad (5)$$

from which relation, together with the obvious fact that

$$J_x|_{z=2\pi nL} = 0, \quad x<0, \quad (6)$$

it is evident that our problem is amenable to the Wiener–Hopf technique.

More specifically, the situation here differs from the usual Wiener–Hopf problem only in the periodicity of the mixed boundary conditions (5) and (6). This difference is conveniently dealt with by noting the symmetry

$$\mathcal{E}_x(z,t) = \mathcal{E}_x(z+2\pi nL, t+2\pi nL/v), \quad (7)$$

which suggests for $\mathcal{E}_x$ the appropriately modified Fourier representation

$$\mathcal{E}_x = \frac{1}{2\pi}\sum_{n=-\infty}^{\infty}\int_{-\infty}^{\infty}d\omega\exp\left[i\omega\left(\frac{z}{v}-t\right)\right]e^{inz/L}\mathcal{E}_n(x,\omega). \quad (8)$$

Here and below, ω should be assumed to have a small positive imaginary part, so that only outgoing waves are obtained. The currents, which possess the same symmetry, (7), as $\mathcal{E}_x$, may be similarly expressed:

$$J_x = 2\pi Lj(x,z-vt)\sum_{n=-\infty}^{\infty}\delta(z-2\pi nL) = \frac{1}{2\pi}\sum_{-\infty}^{\infty}\int_{-\infty}^{\infty}d\omega j(x,\omega)\exp\left[i\omega\left(\frac{z}{v}-t\right)\right]e^{inz/L}, \quad (9)$$

$$J_z = \frac{1}{2\pi}\sum_{-\infty}^{\infty}\int_{-\infty}^{\infty}d\omega q\delta_{n0}\delta(x+x_0) \times\exp\left[i\omega\left(\frac{z}{v}-t\right)\right]e^{inz/L}. \quad (10)$$

Upon substituting these representations into Eq. (3) we find that $\mathcal{E}_n$ and j must satisfy

$$\left[\frac{\partial^2}{\partial x^2}+\omega^2-\left(\frac{\omega}{v}+\frac{n}{L}\right)^2\right]\mathcal{E}_n(x,\omega) = \frac{4\pi}{i\omega}\left(\omega^2+\frac{\partial^2}{\partial x^2}\right)j(x,\omega)+\frac{4\pi q}{v}\delta_{n0}\delta'(x+x_0). \quad (11)$$

The boundary conditions (5) and (6) now take the form

$$\sum_{-\infty}^{\infty}\mathcal{E}_n(x,\omega) = 0, \quad x>0, \quad (12)$$

$$j(x,\omega) = 0, \quad x<0. \quad (13)$$

The system (11)–(13) may be solved by—essentially—the conventional Wiener–Hopf technique.[15] We first Fourier-transform in x, according to the convention

$$\tilde{f}(k) = \int_{-\infty}^{\infty}dx f(x)e^{-ikx},$$

and note that Eqs. (12) and (13) imply analyticity properties for the transformed functions. Thus Eq. (11) becomes

$$(\alpha_n^2+k^2)\tilde{\mathcal{E}}_n(k,\omega) = -\frac{4\pi}{i\omega}(\omega^2-k^2)\tilde{j}_-(k,\omega) - \frac{4\pi q}{v}\delta_{n0}ike^{ikx_0}, \quad (14)$$

where

$$\alpha_n \equiv [(\omega/v+n/L)^2-\omega^2]^{\frac{1}{2}} \quad (15)$$

is defined to have a positive real part. The subscript on $\tilde{j}_-$ serves to remind us that this function must, by Eq. (13), be analytic (in k) in the half-plane Im $(k)<0$. Similarly Eq. (12) implies that the function

$$E_+(k) \equiv \sum_{-\infty}^{\infty}\tilde{\mathcal{E}}_n(k,\omega)$$

is analytic for Im $(k)>0$. But, from Eq. (14),

$$E_+(k) = -\frac{4\pi iq}{v}\frac{ke^{ikx_0}}{k^2+\alpha_0^2} - \frac{4\pi i}{\omega}\tilde{j}_-(k,\omega)(k^2-\omega^2)V(k), \quad (16)$$

where

$$V(k) \equiv \sum_{-\infty}^{\infty}\frac{1}{k^2+\alpha_n^2}, \quad (17)$$

so that both the unknown functions $\tilde{j}_-$ and E_+ may be determined from their analyticity properties, as follows. We suppose there exist functions $V_+(k)$ and $V_-(k)$ such that:

(i) $V_+(k)[V_-(k)]$ is analytic and nonzero in the half-plane Im $(k)\geq 0$ [Im $(k)\leq 0$];

(ii) both the $V_\pm(k)$ have at most polynomial growth for large k;

(iii) $V_+(k)V_-(k) = V(k)$.

We will compute the $V_\pm(k)$ explicitly below; for the present it suffices to note that they have the asymptotic behavior

$$V_\pm(k) \sim k^{-\frac{1}{2}} \quad \text{for} \quad k\to\infty, \quad \text{Im}\,(k)\gtrless 0, \quad (18)$$

and that they evidently allow us to rewrite Eq. (16) in the form

$$\frac{E_+(k)}{V_+(k)}(k - i\alpha_0) + \frac{4\pi iq}{v}\frac{ke^{ikx_0}}{(k + i\alpha_0)}\frac{1}{V_+(k)} = -\frac{4\pi i}{\omega}(k^2 - \omega^2)(k - i\alpha_0)\tilde{j}_-(k)V_-(k). \quad (19)$$

Since the left (right)-hand side of this equation is analytic in the upper (lower)-half k plane, it defines, by analytic continuation, an entire function. That the entire function must be a polynomial—of degree one at most—follows from Eq. (18) and the fact that, for physically acceptable fields and currents, E_+ must vanish for large k. Thus we have, in particular,

$$\tilde{j}_-(k) = \frac{i\omega}{4\pi}\frac{A_0 + A_1k}{(k^2 - \omega^2)(k - i\alpha_0)V_-(k)}. \quad (20)$$

The constants A_0 and A_1 are easily determined. We recall that ω has a small positive imaginary part, so that Eq. (20) is consistent with the analyticity property of $\tilde{j}_-$ only if $A_0 + A_1k = B(k + \omega)$; and the left-hand side of Eq. (19) may be evaluated at $k = i\alpha_0$ to yield

$$B = \frac{2\pi iq}{v}\frac{e^{-\alpha_0x_0}}{(\omega + i\alpha_0)}\frac{1}{V_+(i\alpha_0)}.$$

Finally then, using Eq. (14),

$$\tilde{\mathcal{E}}_n(k, \omega) = -\frac{4\pi iq}{v}\delta_{n0}\frac{ke^{ikx_0}}{k^2 + \alpha_0^2} + \frac{2\pi iq}{v}\frac{e^{-\alpha_0x_0}(k + \omega)}{(\omega + i\alpha_0)V_+(i\alpha_0)(k^2 + \alpha_n^2)(k - i\alpha_0)V_-(k)}. \quad (21)$$

We recognize the first term here as the infinite space solution. Hence the x component of that field, $\mathcal{E}_s$, which arises purely from the surface currents in the plates, is given by Eq. (8) and

$$\mathcal{E}_{sn}(x, \omega) = +\frac{iq}{v}\int_{-\infty}^{\infty}dk \times \frac{e^{-\alpha_0x_0}(k + \omega)e^{ikx}}{(\omega + i\alpha_0)V_+(i\alpha_0)(k^2 + \alpha_n^2)(k - i\alpha_0)V_-(k)}. \quad (22)$$

Note that, for $x < 0$, the integral over k is entirely trivial. In particular,

$$\mathcal{E}_{sn}(-x_0, \omega) = -\frac{\pi q}{v}\frac{e^{-(\alpha_0+\alpha_n)x_0}(\omega - i\alpha_n)}{(\omega + i\alpha_0)V_+(i\alpha_0)\alpha_n(\alpha_0 + \alpha_n)V_-(-i\alpha_n)}. \quad (23)$$

We have now solved Eqs. (3)–(6), except for the determination of V_+ and V_-. We will not examine the field structure here, but will restrict our attention to computing the rate of energy loss, $\dot{W}$, to the plates. This quantity must equal the power needed to move the charged rod through the field due to the plates:

$$\dot{W} = -\int J_z\mathcal{E}_{sz}(x, z, t)\,dx\,dz = -qv\mathcal{E}_{sz}(-x_0, vt, t). \quad (24)$$

Note that at any point not on the plates $\nabla \cdot \mathcal{E}_s = 0$, so that for such points the Fourier components [in the representation of the form (8)] of $\mathcal{E}_{sz}$ may easily be related to those of $\mathcal{E}_{sx}$. In this way Eq. (24) becomes

$$\dot{W} = \frac{qv}{2\pi i}\sum_{n=-\infty}^{\infty}\int_{-\infty}^{\infty}d\omega e^{in\nu t/L}\frac{\alpha_n}{(\omega/v + n/L)}\mathcal{E}_{sn}(-x_0, \omega). \quad (25)$$

Of primary interest is the time-averaged energy loss, to which only the $n = 0$ term in Eq. (25) contributes. We denote the average energy loss per plate by q^2U, that is,

$$U \equiv \frac{2\pi L}{q^2v}\lim_{T\to\infty}\frac{1}{T}\int_0^T \dot{W}\,dt.$$

Equations (23) and (25) give

$$U = -\frac{iLv}{q}\int_{-\infty}^{\infty}d\omega\frac{\alpha_0}{\omega}\mathcal{E}_{s0}(-x_0, \omega) = \frac{\pi iL}{2}\int_{-\infty}^{\infty}\frac{d\omega}{\omega}\frac{e^{-2\alpha_0x_0}}{\alpha_0}\frac{(\omega - i\alpha_0)}{(\omega + i\alpha_0)}\frac{1}{V_+(i\alpha_0)V_-(-i\alpha_0)}. \quad (26)$$

We now turn our attention to the explicit Wiener–Hopf factorization of $V(k)$. This may be accomplished by a conventional procedure.[13,14] We first of all decompose each term of the right-hand side of Eq. (17) into partial fractions

$$[n/L + \omega/L \pm (\omega^2 - k^2)^{\frac{1}{2}}]^{-1}.$$

Using then the identity $\pi\,\mathrm{ctn}\,\pi a = \sum(n + a)^{-1}$, we find

$$V(k) = \frac{\pi L}{2(\omega^2 - k^2)^{\frac{1}{2}}} \times \frac{\sin 2\pi L(\omega^2 - k^2)^{\frac{1}{2}}}{\sin \pi L[\omega/v - (\omega^2 - k^2)^{\frac{1}{2}}]\sin \pi L[\omega/v + (\omega^2 - k^2)^{\frac{1}{2}}]}. \quad (27)$$

But[15]

$$\frac{\sin 2\pi L(\omega^2 - k^2)^{\frac{1}{2}}}{2\pi L(\omega^2 - k^2)^{\frac{1}{2}}} = \prod_{n=1}^{\infty}\left[1 - \frac{4L^2(\omega^2 - k^2)}{n^2}\right] = \prod_{n=1}^{\infty}\left(-i\frac{2L}{n}\right)\left\{k + i\left[\left(\frac{n}{2L}\right)^2 - \omega^2\right]^{\frac{1}{2}}\right\}e^{i2kL/n} \times \prod_{n=1}^{\infty}\left(i\frac{2L}{n}\right)\left\{k - i\left[\left(\frac{n}{2L}\right)^2 - \omega^2\right]^{\frac{1}{2}}\right\}e^{-i2kL/n}, \quad (28)$$

where the $\pm ie^{\mp i2kL/n}$ factors have been inserted so as to make each infinite product converge separately. In a similar manner we obtain

$$2\sin \pi L\left[\frac{\omega}{v} - (\omega^2 - k^2)^{\frac{1}{2}}\right] \sin \pi L\left[\frac{\omega}{v} + (\omega^2 - k^2)^{\frac{1}{2}}\right]$$
$$= 2\pi^2 L^2 (k + i\alpha_0)(k - i\alpha_0)$$
$$\times \prod_{n=1}^{\infty} (-1)\left(\frac{L}{n}\right)^2 (k + i\alpha_n)(k + i\alpha_{-n}) e^{i2kL/n}$$
$$\times \prod_{n=1}^{\infty} (-1)\left(\frac{L}{n}\right)^2 (k - i\alpha_n)(k - i\alpha_{-n}) e^{-i2kL/n}. \quad (29)$$

Upon substituting the representations (28) and (29) into Eq. (27), one may by inspection obtain a factorization

$$V(k) = \hat{V}_+(k)\hat{V}_-(k)$$

in which the factors

$$\hat{V}_+(k) = \frac{1}{(k + i\alpha_0)} \prod_{n=1}^{\infty} \left(\frac{2ni}{L}\right) \frac{\{k + i[(n/2L)^2 - \omega^2]^{\frac{1}{2}}\}}{(k + i\alpha_n)(k + i\alpha_{-n})} \quad (30)$$

and

$$\hat{V}_-(k) = \frac{1}{(k - i\alpha_0)} \prod_{n=1}^{\infty} \left(-\frac{2ni}{L}\right) \frac{\{k - i[(n/2L)^2 - \omega^2]^{\frac{1}{2}}\}}{(k - i\alpha_n)(k - i\alpha_{-n})} \quad (31)$$

clearly have the desired analyticity properties. These factors are, however, unsatisfactory because they are not polynomial bounded for large k.[16] In fact,

$$\hat{V}_\pm(k) \sim 2^{\pm 2ikL} k^{-\frac{1}{2}}, \quad k \to \infty, \quad \mathrm{Im}\,(k) \gtrless 0. \quad (32)$$

[The relation (32) can be seen to follow from Eqs. (30) and (31) by noting the identity

$$\prod_n [1 + (z/n)]e^{-z/n} = [\Gamma(z) z e^{cz}]^{-1}$$

and recalling that $\Gamma(z) \sim e^{-z} z^{z-\frac{1}{2}}$ for large z.[17]] A proper choice for the $V_\pm(k)$ is obviously

$$V_\pm(k) = 2^{\mp 2ikL} \hat{V}_\pm(k), \quad (33)$$

which functions clearly have the asymptotic behavior we anticipated in Eq. (18).

We substitute the results (30)–(33) into Eq. (25), and obtain for the average energy loss per plate (per unit charge squared) the expression

$$U = \frac{2\pi L}{v\gamma} i \int_{-\infty}^{\infty} d\omega \frac{|\omega|}{\omega} \exp\left[-\frac{2|\omega| x_0}{v\gamma}\right] \frac{(\omega - i\alpha_0)}{(\omega + i\alpha_0)}$$
$$\times \left[2^{-2|\omega| L/v\gamma} \prod_{n=1}^{\infty} \frac{(\alpha_0 + \alpha_n)(\alpha_0 + \alpha_{-n})}{\{\alpha_0 + [(n/2L)^2 - \omega^2]^{\frac{1}{2}}\}} \frac{L}{2n}\right]^2.$$

Here γ is the relativistic factor $(1 - v^2)^{-\frac{1}{2}}$ and we have noted, from Eq. (15), that $\alpha_0 = |\omega|/v\gamma$. It is convenient to replace the integration variable by $\lambda \equiv \omega L/v$ and to introduce the abbreviation

$$P(\lambda, \gamma) \equiv \exp\left[-2\frac{|\lambda|}{\gamma} \ln 2\right] \prod_{n=1}^{\infty} \frac{\left\{\frac{|\lambda|}{\gamma} + [(n - \lambda)^2 - \lambda^2 v^2]^{\frac{1}{2}}\right\}\left\{\frac{|\lambda|}{\gamma} + [(n + \lambda)^2 - \lambda^2 v^2]^{\frac{1}{2}}\right\}}{n\left[\frac{2|\lambda|}{\gamma} + (n^2 - 4\lambda^2 v^2)^{\frac{1}{2}}\right]}. \quad (34)$$

Then

$$U = \frac{2\pi i}{\gamma} \int_{-\infty}^{\infty} d\lambda \frac{|\lambda|}{\lambda} \exp\left[-2|\lambda| \frac{x_0}{L\gamma}\right] \left(\frac{\gamma v\lambda - i|\lambda|}{\gamma v\lambda + i|\lambda|}\right) P^2(\lambda, \gamma). \quad (35)$$

Recalling now that ω—and hence λ—is to be considered as having a small positive imaginary part, we find that $P(-\lambda, \gamma) = P^*(\lambda, \gamma)$. This guarantees that U is real and positive, and allows us to rewrite Eq. (35) in the more convenient form

$$U = -\mathrm{Im} \frac{4\pi}{\gamma} \left(\frac{1 - i/v\gamma}{1 + i/v\gamma}\right) \int_0^{\infty} d\lambda$$
$$\times \exp\left[-2\lambda \frac{x_0}{L\gamma}\right] P^2(\lambda, \gamma). \quad (36)$$

The integral in Eq. (36) has been evaluated numerically with the help of Schroeder.[10] The remainder of this report will be devoted to an analytic evaluation of U, for each of the two limiting situations $\gamma \gg 1$ and $\gamma \approx 1$. Specifically, in the case of large γ we will derive a closed form asymptotic expression which is correct to $O(\gamma^{-\frac{3}{2}})$. In the opposite limiting case our expression for U will involve a very easily (but non-analytically) evaluated integral, and will be correct to $O(v^2)$. In both cases a systematic means of obtaining more accurate expressions will be clear, but the labor seems unjustified, especially since our results compare well with the numerical evaluation.

3. ENERGY LOSS FROM AN ULTRARELATIVISTIC ROD

Considering first the case of large γ, we begin with the observation that

$$P(\lambda, \infty) = 1, \quad (37)$$

as is evident from the definition, Eq. (34). Thus our

procedure will be to let

$$z(\lambda, \gamma) \equiv 2 \ln P(\lambda, \gamma), \tag{38}$$

and expand

$$P^2 = 1 + z + z^2/2! + z^3/3! + \cdots. \tag{39}$$

Our first task, then, is to derive a sufficiently accurate expression for z. The definition, Eq. (38), may be substantially simplified if we use Eq. (37) to write

$$P(\lambda, \gamma) = P(\lambda, \gamma)/P(\lambda, \infty). \tag{40}$$

One term in the logarithm of Eq. (38) is

$$\ln\left\{\frac{1}{\gamma} + [(n/\lambda)^2 - 2(n/\lambda) + 1/\gamma^2]^{\frac{1}{2}}\right\} - \ln[(n/\lambda)^2 - 2(n/\lambda)]^{\frac{1}{2}} = \sinh^{-1}\gamma[(n/\lambda)^2 - 2(n/\lambda)]^{-\frac{1}{2}},$$

and the other terms may be treated similarly. We find

$$z(\lambda) = -4\frac{\lambda}{\gamma}\ln 2 + 2\sum_{n=1}^{\infty} F\left(\frac{n}{\lambda}\right), \tag{41}$$

where

$$F(\phi) \equiv \sinh^{-1}\frac{1}{\gamma(\phi^2 - 2\phi)^{\frac{1}{2}}} + \sinh^{-1}\frac{1}{\gamma(\phi^2 + 2\phi)^{\frac{1}{2}}} - \sinh^{-1}\frac{2}{\gamma(\phi^2 - 4)^{\frac{1}{2}}}. \tag{42}$$

Because of Eq. (36) we may restrict our attention to Re $(\lambda) > 0$. Recalling our convention Im $(\lambda) > 0$, we see that the relevant singularities of $F(\phi)$ occur in the upper-half $\phi = (n/\lambda)$ plane, at

$$\phi = 1 \pm (1 - 1/\gamma^2)^{\frac{1}{2}} + i0,$$
$$\phi = 2(1 - 1/\gamma^2)^{\frac{1}{2}} + i0,$$
$$\phi = i0,$$
$$\phi = 2 + i0.$$

These are all branch points; we choose the cuts of $F(\phi)$ for Re $(\phi) > 0$ to extend upwards as indicated in Fig. 2. We consider next

$$\int_0^{\infty} dn F\left(\frac{n}{\lambda}\right)$$
$$= \lambda \lim_{R\to\infty} \int_0^R d\phi F(\phi)$$
$$= \lambda \lim_{R\to\infty}\left\{\int_{-1}^{R-1} + \int_1^{R+1} - 2\int_0^{R/2}\right\} \sinh^{-1}\frac{1}{\gamma(x^2 - 1)^{\frac{1}{2}}}\,dx$$
$$= 2\lambda \lim_{R\to\infty} \int_{R/2}^{R} \sinh^{-1}\frac{1}{\gamma(x^2 - 1)^{\frac{1}{2}}}\,dx.$$

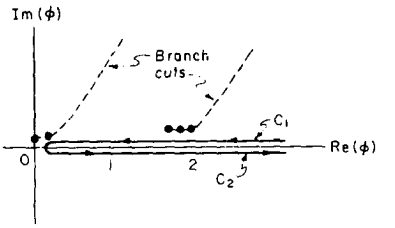

FIG. 2. Branch cuts and integration contours in the complex ϕ plane.

Since $x \in (R/2, R)$ is arbitrarily large, we may replace the integrand by $\gamma^{-1}(x^2 - 1)^{-\frac{1}{2}}$, whence

$$\int_0^{\infty} dn F\left(\frac{n}{\lambda}\right) = \frac{2\lambda}{\gamma}\ln 2. \tag{43}$$

Using also the relation

$$2\sum_{n=1}^{\infty} F\left(\frac{n}{\lambda}\right) = \frac{\lambda}{i}\int_C d\phi F(\phi)\cot \pi\lambda\phi,$$

in which the path $C = C_1 + C_2$ remains below the branch points of F as in Fig. 2, we find

$$z(\lambda) = 2\left[\sum_{n=1}^{\infty} - \int_0^{\infty} dn\right] F\left(\frac{n}{\lambda}\right) \tag{44}$$
$$= \frac{\lambda}{i}\int_{C_1} d\phi F(\phi)(\cot \pi\lambda\phi + i) + \frac{\lambda}{i}\int_{C_2} d\phi F(\phi)(\cot \pi\lambda\phi - i). \tag{45}$$

We now allow C_1 and C_2 to coalesce onto the real axis from the origin up to some point δ, where we choose $\delta \geq 1/\gamma^2$. We evidently omit in this way some residue contributions from the poles of $\cot \pi\lambda\phi$; however, by our choice of δ, such contributions can affect the behavior of $z(\lambda)$ only for very large λ (specifically, $\lambda = n\gamma^2$, $n = 1, 2, \cdots$). It follows that the omitted contributions to z will be weighted in our formula (36) for U by $e^{-\gamma}$, and may clearly be neglected in the present large-γ analysis. Next, we take advantage of the fact that $(\cot \pi\lambda\phi \pm i)$ becomes exponentially small in the limit $\phi \to \pm i\infty$, by deforming the rest of the contour C_2 into the lower-half plane and by "wrapping" the contour C_1 around the branch cuts in the usual way. The result of these mutilations is to leave Eq. (45) in the form

$$z(\lambda) \cong \frac{\lambda}{i}\int_{\mathrm{I}+\mathrm{II}} (\cot \pi\lambda\phi + i)F(\phi)\,d\phi + \frac{\lambda}{i}\int_{\mathrm{III}} (\cot \pi\lambda\phi - i)F(\phi)\,d\phi + 2\lambda\int_{\mathrm{IV}} F(\phi)\,d\phi,$$

which is conveniently decomposed as follows:

$$z(\lambda) = z_1(\lambda) + z_2(\lambda) + z_3(\lambda), \tag{46}$$

$$z_1(\lambda) \equiv \frac{\lambda}{i}\int_{\mathrm{I}} (\cot \pi\lambda\phi + i)F(\phi)\, d\phi, \tag{47}$$

$$\begin{aligned} z_2(\lambda) \equiv & -\frac{\lambda}{i}\int_{\mathrm{II}} (\cot \pi\lambda\phi + i)\sinh^{-1}\frac{2}{\gamma(\phi^2-4)^{\frac{1}{2}}}\, d\phi \\ & -\frac{\lambda}{i}\int_{\mathrm{III}} (\cot \pi\lambda\phi - i)\sinh^{-1}\frac{2}{\gamma(\phi^2-4)^{\frac{1}{2}}}\, d\phi \\ & -2\lambda\int_{\mathrm{IV}} \sinh^{-1}\frac{2}{\gamma(\phi^2-4)^{\frac{1}{2}}}\, d\phi, \end{aligned} \tag{48}$$

$$\begin{aligned} z_3(\lambda) \equiv & \frac{\lambda}{i}\int_{\mathrm{II}} (\cot \pi\lambda\phi + i)f(\phi)\, d\phi \\ & +\frac{\lambda}{i}\int_{\mathrm{III}} (\cot \pi\lambda\phi - i)f(\phi)\, d\phi \\ & +2\lambda\int_{\mathrm{IV}} f(\phi)\, d\phi, \end{aligned} \tag{49}$$

where

$$f(\phi) \equiv \sinh^{-1}\frac{1}{\gamma(\phi^2-2\phi)^{\frac{1}{2}}} + \sinh^{-1}\frac{1}{\gamma(\phi^2+2\phi)^{\frac{1}{2}}}. \tag{50}$$

The contours I–IV are depicted in Fig. 3.

We re-emphasize that the error in Eq. (46) corresponds to an error in U which is exponentially small; up to this point, no serious approximations have been made. Now, however, we commit ourselves to keeping only $O(\gamma^{-\frac{3}{2}})$ terms in Eq. (36), which then becomes

$$\begin{aligned} U = & -\mathrm{Im}\frac{4\pi}{\gamma}\left(1 - \frac{2i}{v\gamma}\right)\int_0^\infty d\lambda \\ & \times \exp\left[-2\frac{x_0}{L\gamma}\lambda\right]\left(1 + z + \frac{z^2}{2!} + \frac{z^3}{3!}\right) + O(\gamma^{-2}). \end{aligned} \tag{51}$$

[We anticipate here that each power of z contributes $O(\gamma^{-\frac{1}{2}})$ to U.] The contribution from z_1 to Eq. (51) is easily disposed of. Consider

$$\begin{aligned} U_1 \equiv & -\mathrm{Im}\frac{4\pi}{\gamma}\left(1 - \frac{2i}{\gamma}\right)\int_0^\infty d\lambda \exp\left[-2\frac{x_0}{L\gamma}\lambda\right] z_1 \\ & \propto \frac{1}{\gamma^2}\int_{\mathrm{I}} F(\phi)\sum_{m=1}^{\infty}\frac{d\phi}{(2x_0/L\gamma - 2\pi i m\phi)^2}, \end{aligned} \tag{52}$$

where we have used the identity

$$\cot\frac{x}{2} = \frac{2}{i}\sum_{n=1}^{\infty} e^{inx} - i - \frac{2\pi}{i}\sum_{-\infty}^{\infty}\delta(x - 2\pi n), \tag{53}$$

in which the δ-function terms are here irrelevant; to explicitly perform the integral over λ. Since $|\phi|$ is not small on the path I, it is clear that

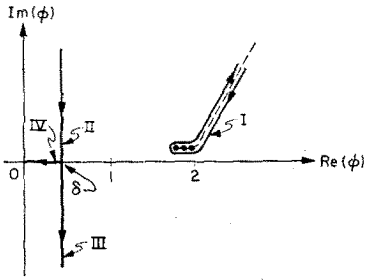

FIG. 3. Transformed integration contours.

$$U_1 = O(\gamma^{-2}).$$

The higher-order terms in z_1 may be similarly treated, and we conclude that, in our approximation, the term z_1 may be omitted from Eq. (46).

With regard to z_2, we observe first of all that the definition (48) yields, by nature of the paths II, III, IV,

$$\begin{aligned} z_2 = & -\frac{\lambda}{i}\int_{\mathrm{II+III}} \cot \pi\lambda\phi \sinh^{-1}\frac{2}{\gamma(\phi^2-4)^{\frac{1}{2}}}\, d\phi \\ & -\lambda\int_{i\infty}^{\delta} \sinh^{-1}\frac{2}{\gamma(\phi^2-4)^{\frac{1}{2}}}\, d\phi \\ & +\lambda\int_{\delta}^{-i\infty} \sinh^{-1}\frac{2}{\gamma(\phi^2-4)^{\frac{1}{2}}}\, d\phi \\ & -2\lambda\int_{\delta}^{0} \sinh^{-1}\frac{2}{\gamma(\phi^2-4)^{\frac{1}{2}}}\, d\phi. \end{aligned} \tag{54a}$$

By changing the sign of the integration variable in the second term of this expression, we see that the last three terms cancel. In the remaining first term, we again ignore residue contributions from the poles of $\cot \pi\lambda\phi$ for $0 < \phi < \delta$ [cf. the remarks following Eq. (45)], so that the integration contour II + III may be considered as displaced infinitesimally to the right of the imaginary ϕ axis. Hence we may evaluate this term,

$$z_2 = -\frac{\lambda}{i}\int_{i\infty+\delta}^{-i\infty+\delta} \cot \pi\lambda\phi \sinh^{-1}\frac{2}{\gamma(\phi^2-4)^{\frac{1}{2}}}\, d\phi, \tag{54b}$$

in terms of principal value and pole (at $\phi = 0$) contributions in the usual way. Since the integrand is odd in ϕ, the principal value vanishes and we are left with the semiresidue

$$\begin{aligned} z_2 &= \sinh^{-1}(i/\gamma) \\ &= i/\gamma + O(\gamma^{-3}). \end{aligned} \tag{55}$$

The calculation of z_3 [Eq. (49)] is more complicated. It is helpful to observe first, from Eq. (51), that a term in z_3 of the form $\lambda^\alpha\gamma^\beta$ can contribute to U at most a term of order $\gamma^{\alpha+\beta}$ (its contribution will be smaller if the term is purely real). Thus we may drop such terms

whenever $\alpha + \beta < -\frac{3}{2}$. With this in mind we expand Eq. (50):

$$f(\phi) = f^0(\phi) + f^1(\phi) + O(\phi^{\frac{3}{2}}), \tag{56}$$

where

$$f^0(\phi) \equiv \sinh^{-1}\frac{1}{\gamma(2\phi)^{\frac{1}{2}}} + \sinh^{-1}\frac{1}{\gamma(-2\phi)^{\frac{1}{2}}}, \tag{57}$$

$$f^1(\phi) \equiv \frac{(i-1)}{4\sqrt{2}}\frac{\phi^{\frac{1}{2}}}{\gamma}. \tag{58}$$

The error in Eq. (56) would contribute to z_3 a term of order $\lambda^{-\frac{3}{2}}\gamma^{-1}$ and, by our remarks above, may be ignored. To compute the contribution to z_3 from f^0, it is simplest to revert from Eq. (49) to our earlier formulation, Eq. (44):

$$\begin{aligned} z_3^0 &= 2\left[\sum_{n=1}^{\infty} - \int_0^{\infty} dn\right] f^0\left(\frac{n}{\lambda}\right) \\ &= 2\left[\sum_{n=1}^{\infty} - \int_0^{\infty} dn\right]\frac{(1+i)}{\gamma(2n/\lambda)^{\frac{1}{2}}} \\ &\quad + 2\left[\sum_{n=1}^{\infty} - \int_0^{\infty} dn\right]\left[f^0\left(\frac{n}{\lambda}\right) - \frac{(1+i)}{\gamma(2n/\lambda)^{\frac{1}{2}}}\right]. \end{aligned} \tag{59a}$$

With regard to the first term here we note that

$$\left[\sum_{n=1}^{\infty} - \int_0^{\infty} dn\right] n^{-\frac{1}{2}} = \zeta(\tfrac{1}{2}), \tag{59b}$$

where ζ is the Riemann zeta function.[18] For the *sum* in the second term, we expand $\sinh^{-1} x - x = (x^3/6) + O(x^5)$. The error term may again be ignored, whence

$$\sum_{n=1}^{\infty}\left[f^0\left(\frac{n}{\lambda}\right) - \frac{(1+i)}{\gamma(2n/\lambda)^{\frac{1}{2}}}\right] = \frac{(1-i)(\lambda/2)^{\frac{3}{2}}}{6\gamma^3}\zeta(\tfrac{3}{2}). \tag{59c}$$

Finally, the *integral* in the last term of Eq. (59a),

$$\begin{aligned} &\int_0^{\infty} dn\left[f^0\left(\frac{n}{\lambda}\right) - \frac{(1+i)}{\gamma(2n/\lambda)^{\frac{1}{2}}}\right] \\ &\quad = \frac{\lambda}{2\gamma^2}\int_{-\infty}^{\infty} dx x^{-3}[\sinh^{-1} x + \sinh^{-1} ix - (1+i)x], \end{aligned} \tag{59d}$$

vanishes identically, as may be verified by integrating by parts twice. We have then

$$z_3^0 = (\sqrt{2})(1+i)\zeta(\tfrac{1}{2})\frac{\lambda^{\frac{1}{2}}}{\gamma} - \frac{(1-i)}{6\sqrt{2}}\zeta(\tfrac{3}{2})\frac{\lambda^{\frac{3}{2}}}{\gamma^3}. \tag{60}$$

Next we must compute the contribution to z_3 of f^1. According to Eqs. (49) and (58), this is

$$\begin{aligned} z_3^1 = \frac{(1-i)}{4\sqrt{2}}\, i\frac{\lambda}{\gamma}\Bigg\{&\int_{\mathrm{II}}(\cot \pi\lambda\phi + i)\phi^{\frac{1}{2}}\, d\phi \\ &+ \int_{\mathrm{III}}(\cot \pi\lambda\phi - i)\phi^{\frac{1}{2}}\, d\phi + 2i\int_{\mathrm{IV}}\phi^{\frac{1}{2}}\, d\phi\Bigg\}. \end{aligned} \tag{61}$$

By the nature of the path IV the last term here is $O(\lambda/\gamma^3)$ and may be neglected; the other terms may be integrated explicitly by means of Eq. (53) (in which again the δ functions do not contribute). We have then

$$\begin{aligned} z_3^1 &= \frac{(1-i)\lambda}{2(\sqrt{2})\gamma}\left\{\int_{i\infty}^{0}\sum_{n=1}^{\infty} e^{2\pi i\lambda n\phi}\phi^{\frac{1}{2}}\, d\phi \right. \\ &\quad \left. - \int_0^{-i\infty}\sum_{n=1}^{\infty} e^{-2\pi i\lambda n\phi}\phi^{\frac{1}{2}}\, d\phi\right\} \\ &= \frac{(1-i)\lambda}{2\gamma}\int_0^{\infty}\sum_{n=1}^{\infty} e^{-2\pi\lambda n y}y^{\frac{1}{2}}\, dy \\ &= \frac{(1-i)}{8(\sqrt{2})\pi}\frac{\lambda^{-\frac{1}{2}}}{\gamma}\zeta(\tfrac{3}{2}). \end{aligned} \tag{62}$$

Equations (60) and (62) provide a sufficiently accurate expression for $z_3 = z_3^0 + z_3^1$. This result, together with Eq. (55), is now, of course, to be substituted into Eq. (51). Thus, after performing the elementary integration over λ, we find that the average energy loss per plate is

$$\begin{aligned} q^2U = &-q^2\frac{\zeta(\frac{1}{2})}{2\sqrt{2}}\left(\frac{2\pi L}{x_0}\right)^{\frac{3}{2}}\gamma^{-\frac{1}{2}} \\ &+ q^2\left[1 - \frac{\zeta^2(\frac{1}{2})}{2\pi}\left(\frac{2\pi L}{x_0}\right)\right]\left(\frac{2\pi L}{x_0}\right)\gamma^{-1} \\ &+ q^2\left\{\frac{\zeta(\frac{3}{2})}{4\sqrt{2}} + \frac{\zeta(\frac{1}{2})}{2\sqrt{2}}\left(\frac{2\pi L}{x_0}\right)\right. \\ &\left. - \frac{1}{8(\sqrt{2})\pi}\left[\zeta^3(\tfrac{1}{2}) + \frac{\zeta(\frac{3}{2})}{8}\right]\left(\frac{2\pi L}{x_0}\right)^2\right\}\left(\frac{2\pi L}{x_0}\right)^{\frac{1}{2}}\gamma^{-\frac{3}{2}} \\ &+ O(\gamma^{-2}). \end{aligned} \tag{63}$$

Recall that $2\pi L$ is the actual distance between plates. We introduce the abbreviation

$$\rho \equiv 2\pi L/x_0,$$

and evaluate[19] the numerical coefficients in Eq. (63), which then becomes

$$U = a\gamma^{-\frac{1}{2}} + b\gamma^{-1} + c\gamma^{-\frac{3}{2}} + O(\gamma^{-2}), \tag{64}$$

with

$$a \cong 0.516\rho^{\frac{3}{2}}, \tag{65}$$

$$b \cong \rho(1 - 0.339\rho), \tag{66}$$

$$c \cong \rho^{\frac{1}{2}}(0.462 - 0.516\rho + 0.0784\rho^2). \tag{67}$$

In Fig. 4 evaluation of Eq. (64)—for three values of ρ—is presented and compared with numerical evaluation of Eq. (36). The accuracy of the asymptotic formulas—even to γ as low as 2.0—was, of course, unexpected.

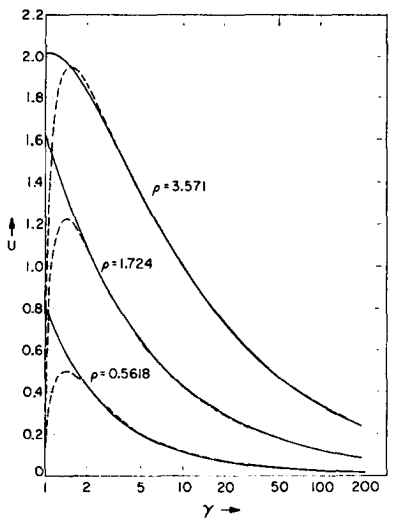

FIG. 4. Energy loss per plate, per square of the unit charge per unit length on the rod, as a function of rod speed v, expressed in terms of the relativistic factor

$$\gamma \equiv (1 - v^2)^{-\frac{1}{2}}.$$

The solid lines are the asymptotic evaluation of Eq. (64) and the dashed lines are the numerical evaluation of Eq. (36). Curves are presented—as indicated on the figure—for three values of

$$\rho \equiv 2\pi L/x_0$$

which span the range of practical interest.

4. ENERGY LOSS FROM A SLOW ROD

We now turn our attention to the case in which γ is close to 1. It is convenient to begin again with our exact formula (36), but here we will always drop terms of order v^3. The small velocity limit is analytically somewhat awkward because of the form of lim Re (P). Equation (34) gives

$$P(\lambda, 1) = e^{-2\lambda \ln 2} \prod_{n=1}^{[\lambda]} \frac{2\lambda - n}{n}, \tag{68}$$

where $[\lambda]$ is the largest integer less than λ. On the other hand we will find that

$$\text{Im}\,[P^2(\lambda, \gamma)] = O(v^2), \quad v \approx 0, \tag{69}$$

from which it follows in particular that we may approximate

$$\text{Im}\,(P^2) \cong 2P^2(\lambda, 1)\,\text{Im}\,(\ln P). \tag{70}$$

Similarly expanding

$$-\frac{1}{\gamma}\frac{1 - i/v\gamma}{1 + i/v\gamma} = 1 + 2iv - \tfrac{5}{2}v^2 + O(v^3),$$

where the v^2 term is irrelevant in view of Eq. (69), we find that Eq. (36) may be written in the form

$$U = 8\pi(vU_0 + U_1) + O(v^3), \tag{71}$$

$$U_0 \equiv \int_0^\infty d\lambda e^{-\kappa\lambda} P^2(\lambda, 1), \tag{72}$$

$$U_1 \equiv \int_0^\infty d\lambda e^{-\kappa\lambda} P^2(\lambda, 1)\,\text{Im}\,(\ln P). \tag{73}$$

Here, of course, $\kappa = 2x_0/L$, and we need evaluate Im $(\ln P)$ only to lowest order.

We consider first U_0. The definition (72) may be rewritten as

$$U_0 = \sum_{m=0}^{\infty} \int_0^1 dx e^{-\kappa(m+x)} P^2(m + x, 1). \tag{74}$$

From Eq. (68),

$$P(m + x, 1) = e^{-2(m+x)\ln 2}\binom{2m + 2x - 1}{m}, \tag{75}$$

where the second factor is the usual binomial coefficient. Introducing the quantity

$$z \equiv \tfrac{1}{4}e^{-\kappa/2},$$

we have

$$U_0 = \int_0^1 dx z^{2x} \sum_{m=0}^{\infty}\left[z^m\binom{2m + 2x - 1}{m}\right]^2. \tag{76}$$

The sum of squares may be rewritten as the square of a sum by means of the artifice

$$\sum_{m=0}^{\infty}\left[z^m\binom{2m + 2x - 1}{m}\right]^2 = \frac{1}{2\pi}\int_0^{2\pi} d\theta \left|\sum_{m=0}^{\infty}(e^{i\theta}z)^m\binom{2m + 2x - 1}{m}\right|^2. \tag{77}$$

We denote the sum on the right-hand side by $S(x, \theta)$, so that

$$U_0 = \int_0^1 dx z^{2x}\frac{1}{2\pi}\int_0^{2\pi} d\theta\,|S(x, \theta)|^2. \tag{78}$$

By the binomial theorem and Cauchy's theorem,

$$S(x, \theta) = \frac{1}{2\pi i}\sum_{m=0}^{\infty}(ze^{i\theta})^{2m}\oint \frac{dw}{w^{m+1}}\left(1 + \frac{w}{ze^{i\theta}}\right)^{2m+2x-1}, \tag{79}$$

where the integration contour must enclose the origin of the w plane in such a way as to include only the pole at $w = 0$. We choose it to be a circle with radius only slightly less than $\frac{1}{4}$. The series is now geometric and easily summed:

$$S(x, \theta) = \frac{1}{2\pi i}\oint \frac{dw}{w}\left(1 + \frac{w}{ze^{i\theta}}\right)^{2x-1}\frac{w}{w - (w + ze^{i\theta})^2}. \tag{80}$$

The integrand here has poles at

$$w = \tfrac{1}{2} - ze^{i\theta} \pm \tfrac{1}{2}(1 - 4ze^{i\theta})^{\frac{1}{2}}$$

and by our contour choice only the smaller of these is enclosed. Since the integrand is otherwise analytic inside the contour,

$$S(x, \theta) = (ze^{i\theta})^{1-2x}\frac{[\frac{1}{2} - \frac{1}{2}(1 - 4ze^{i\theta})^{\frac{1}{2}}]^{2x-1}}{(1 - 4ze^{i\theta})^{\frac{1}{2}}}. \tag{81}$$

It is now convenient to replace z by

$$u \equiv 4z = e^{-\kappa/2}$$

so that upon substituting Eq. (81) into Eq. (78) we have

$$U_0 = \frac{1}{4}\int_0^1 dx u^{2x}\frac{1}{2\pi}\int_0^{2\pi} d\theta u^{2-4x} \times \frac{[1-(1-ue^{i\theta})^{\frac{1}{2}}]^{2x-1}[1-(1-ue^{-i\theta})^{\frac{1}{2}}]^{2x-1}}{(1-ue^{i\theta})^{\frac{1}{2}}(1-ue^{-i\theta})^{\frac{1}{2}}}. \tag{82}$$

Notice that x appears now only in exponents, so that the x integration could readily be performed. However, we defer this step in order first to simplify the integral over θ. To this end we replace the integration variable by $y = e^{i\theta}$, whence

$$U_0 = \frac{1}{4}\int_0^1 dx u^{2(1-x)}\frac{1}{2\pi i}\oint\frac{dy}{y} \times \frac{\{[1-(1-uy)^{\frac{1}{2}}][1-(1-u/y)^{\frac{1}{2}}]\}^{2x-1}}{(1-uy)^{\frac{1}{2}}(1-u/y)^{\frac{1}{2}}}. \tag{83}$$

Here the contour is the unit circle. The integrand has branch points at $y = 0, u, 1/u, \infty$. Since it is evident from Eq. (82) that the integrand is continuous throughout the domain of integration, no branch cut can cross the contour. Thus the branch points at 0 and u must be connected by a cut, and we shrink the contour to surround this cut in the obvious way. On the new contour y is real and $0 \leq y \leq u$ so that $(1-u/y)^{\frac{1}{2}}$ is always imaginary. We thus make the further substitution

$$(1-u/y)^{\frac{1}{2}} = i\tan\theta,$$

and obtain

$$U_0 = \frac{1}{4\pi}\int_0^1 dx u^{2(1-x)}\int_{-\pi/2}^{\pi/2} d\theta \times \frac{\{[1-(1-u^2\cos^2\theta)^{\frac{1}{2}}](1-i\tan\theta)\}^{2x-1}}{(1-u^2\cos^2\theta)^{\frac{1}{2}}}$$
$$= \frac{1}{4\pi}\int_{-\pi/2}^{\pi/2} d\theta \frac{u^2 e^{i\theta}\cos\theta}{(1-u^2\cos^2\theta)^{\frac{1}{2}}[1-(1-u^2\cos^2\theta)^{\frac{1}{2}}]} \times \int_0^1 dx\left[\frac{1-(1-u^2\cos^2\theta)^{\frac{1}{2}}}{e^{i\theta}u\cos\theta}\right]^{2x}. \tag{84}$$

Performing the x integral we find, after some simple manipulations,

$$U_0 = \frac{1}{4\pi}\int_{-\pi/2}^{\pi/2} d\theta \frac{(1-u^2\cos^2\theta)^{\frac{1}{2}} - i\tan\theta}{(1-u^2\cos^2\theta)^{\frac{1}{2}}} \times \frac{1}{\cosh^{-1}\dfrac{1}{u\cos\theta} - i\theta}. \tag{85}$$

This form may be further simplified if we observe that

$$\frac{-i\tan\theta}{(1-u^2\cos^2\theta)^{\frac{1}{2}}} = -i\frac{\partial}{\partial\theta}\left(\cosh^{-1}\frac{1}{u\cos\theta} - i\theta\right) + 1.$$

Hence the $\tan\theta$ term in Eq. (85) may be integrated by parts; we thus find its contribution is equal to that of the other term. We have then, finally,

$$U_0 = \frac{1}{2\pi}\int_{-\pi/2}^{\pi/2} d\theta \Big/ \left(\cosh^{-1}\frac{1}{u\cos\theta} - i\theta\right)$$
$$= \frac{1}{\pi}\int_0^{\pi/2} d\theta \left(\cosh^{-1}\frac{1}{u\cos\theta}\right) \Big/ \left[\left(\cosh^{-1}\frac{1}{u\cos\theta}\right)^2 + \theta^2\right]. \tag{86}$$

This integral is easily evaluated numerically; a plot of $U_0(u)$ is given in Fig. 5. It has been checked—approximately—by extrapolations of numerical evaluations of Eq. (36).

The asymptotic forms of U_0 for large and small $\kappa = 2x_0/L$ are easily determined. Considering first the case of large κ, we note that $\cosh^{-1}(u\cos\theta)^{-1} \approx -\ln 2u\cos\theta$, for $u \approx 0$. Hence in this limit

$$U_0 \approx -\tfrac{1}{2}(\ln u)^{-1} = L/2x_0, \quad x_0 \gg L. \tag{87}$$

On the other hand, for κ small and $\theta \leqslant \pi/4$,

$$\cosh^{-1}(u\cos\theta)^{-1} \approx (\kappa+\theta^2)^{\frac{1}{2}} \quad \text{for} \quad \kappa \approx 0, \theta \leqslant \pi/4,$$

so that the main contribution to the integral in Eq. (86) comes from the lower end point. It is in fact clear that if we choose δ to be proportional to, but larger than, $\kappa^{\frac{1}{2}}$, the small κ form of U_0 will follow the small δ form of

$$\frac{1}{\pi}\int_\delta^{\pi/2}\cosh^{-1}\frac{1}{u\cos\theta}\Big/\left[\left(\cosh^{-1}\frac{1}{u\cos\theta}\right)^2+\theta^2\right]d\theta \approx \frac{1}{\pi}\int_\delta^{\pi/2}\left\{\ln\frac{1+\sin\theta}{\cos\theta}\Big/\left[\left(\ln\frac{1+\sin\theta}{\cos\theta}\right)^2+\theta^2\right] - \frac{1}{2\theta}\right\}d\theta + \frac{1}{2\pi}\int_\delta^{\pi/2}\frac{d\theta}{\theta}. \tag{88}$$

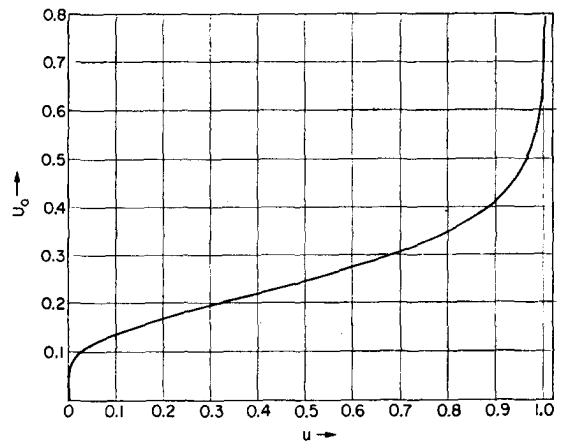

FIG. 5. The integral U_0, defined in Eq. (86), as a function of $u = e^{-x_0/L}$. To lowest order in rod speed, v, the energy loss per plate, per square of the unit charge per unit length on the rod, is given by $U = 8\pi v U_0$.

Here we have set $u = 1$ in the integrand (since $\theta^2 \geq \delta^2 > \kappa$) and isolated the singular part of U_0 ($u = 1$). Now letting $\delta \propto \kappa^{\frac{1}{2}}$ become arbitrarily small, we find

$$U_0 \approx -\frac{1}{2\pi} \ln \delta, \qquad \delta \approx 0,$$

$$\approx -\frac{1}{4\pi} \ln \frac{x_0}{L}, \quad x_0 \ll L, \tag{89}$$

since the first integral in Eq. (88) remains finite as $\delta \to 0$.

In order to evaluate U_1 [cf. Eq. (73)], we first note from Eq. (34) that for small v

$$\text{Im}\,(\ln P) = \text{Im} \sum_{n=1}^{\infty} \ln \left\{ \frac{\lambda}{\gamma} + [(n-\lambda)^2 - \lambda^2 v^2]^{\frac{1}{2}} \right\} \tag{90}$$

$$\approx \text{Im} \sum_{n=1}^{\infty} \ln \left\{ 1 + \left[\left(\frac{n}{\lambda} - 1 \right)^2 - v^2 \right]^{\frac{1}{2}} \right\} \tag{91}$$

since, as we have remarked, it suffices to compute U_1 to lowest order. We now set $\lambda = m + x$, but here, unlike the case of Eq. (74), we choose $|x| \leq \frac{1}{2}$. Then

$$\text{Im}\,(\ln P) \approx \sum_{n=1}^{\infty} \text{Im} \ln \left\{ 1 + i \left[v^2 - \left(\frac{n}{m+x} - 1 \right)^2 \right]^{\frac{1}{2}} \right\} \tag{92}$$

$$\approx \Theta\left(\frac{mv}{1-v} - |x| \right) \left[v^2 - \left(\frac{x}{m+x} \right)^2 \right]^{\frac{1}{2}}, \tag{93}$$

where Θ is the usual Heaviside function; it is equal to one (zero) when its argument is positive (negative). Equation (73) has become

$$U_1 = \sum_{m=1}^{\infty} \int_{-\eta}^{\eta} dx e^{-\kappa(m+x)} P^2(m+x, 1) \times \left[v^2 - \left(\frac{x}{m+x} \right)^2 \right]^{\frac{1}{2}}, \tag{94}$$

where $\eta \equiv mv/(1-v)$. Because of the exponential factor in the integrand, we may assume η is small. It follows that to lowest order

$$U_1 = \sum_{m=1}^{\infty} e^{-\kappa m} P^2(m, 1) \int_{-\eta}^{\eta} \left[v^2 - \left(\frac{x}{m+x} \right)^2 \right]^{\frac{1}{2}} dx. \tag{95}$$

The integral is $(\pi/2)mv^2 + O(v^3)$, whence

$$U_1 = -\frac{\pi}{2} v^2 \frac{\partial}{\partial \kappa} \sum_{m=0}^{\infty} e^{-\kappa m} P^2(m, 1). \tag{96}$$

Here we have included the $m = 0$ term—which clearly does not contribute—so that the sum may be recognized from Eq. (79):

$$\sum_{m=0}^{\infty} e^{-\kappa m} P^2(m, 1) = \frac{1}{2\pi} \int_0^{2\pi} d\theta\, |S(0, \theta)|^2.$$

We have thus merely to repeat the procedure of Eqs. (74)–(84) above, setting everywhere $x = 0$. The result is

$$U_1 = \frac{-v^2}{8} \frac{\partial}{\partial \kappa} \int_{-\pi/2}^{\pi/2} \frac{d\theta}{(1 - u^2 \cos^2 \theta)^{\frac{1}{2}}}$$

$$= \frac{-v^2}{4} \frac{\partial}{\partial \kappa} K(e^{-\kappa/2}), \tag{97}$$

where K is the complete elliptic integral of the first kind. The identity[19]

$$\frac{dK}{dk} = \frac{1}{k} \left(\frac{E(k)}{1-k^2} - K(k) \right),$$

in which E is the complete elliptic integral of the second kind, finally gives

$$U_1 = \frac{v^2}{8} \left\{ \frac{E(e^{-\kappa/2})}{1 - e^{-\kappa}} - K(e^{-\kappa/2}) \right\}. \tag{98}$$

Using known[19] properties of the functions E and K, we find that U_1 vanishes exponentially for large κ, while

$$U_1 \approx \frac{v^2}{8} \frac{1}{\kappa} \quad \text{for} \quad \kappa \approx 0. \tag{99}$$

Combining now Eqs. (71), (86), and (98), we conclude

$$U = 8v \int_0^{\pi/2} d\theta \cosh^{-1} \left(\frac{e^{\kappa/2}}{\cos\theta} \right) \Big/ \left\{ \left[\cosh^{-1} \left(\frac{e^{\kappa/2}}{\cos\theta} \right) \right]^2 + \theta^2 \right\} + \pi v^2 \left[\frac{E(e^{-\kappa/2})}{1 - e^{-\kappa}} - K(e^{-\kappa/2}) \right] + O(v^3). \tag{100}$$

ACKNOWLEDGMENTS

We are very grateful to E P. Lee for correcting numerous mistakes in the original manuscript and to Esther Schroeder for performing the numerical evaluations.

APPENDIX A: THE CASE OF A MOVING CURRENT

Our analysis is very easily carried over to the solution of a slightly different problem, namely, that in which the moving rod has no net charge, but carries a current in the y direction. In this appendix we briefly outline the necessary modifications to the arguments and results presented above.

The essential difference between the two problems is that, in the moving current case, the relevant component of the field is $\mathcal{E}_y$, and this function satisfies an equation significantly simpler than that [Eq. (3)] for $\mathcal{E}_x$ in the moving charge case. In fact we have now to solve

$$\left(\frac{\partial^2}{\partial x^2}+\frac{\partial^2}{\partial z^2}-\frac{\partial^2}{\partial t^2}\right)\mathcal{E}_y = 4\pi\frac{\partial J_y}{\partial t}, \tag{A1}$$

where $\mathcal{E}_y$ satisfies the usual boundary condition (5) and

$$J_y = qv'\delta(x+x_0)\delta(z-vt)+J_{sy}. \tag{A2}$$

Here x_0 and v are as in Fig. 1, and v' is the y-directed velocity, associated with the given current, of the charge per unit length q. [The q in Eq. (A2) has the same numerical value as that in Eq. (4), insofar as the positive charge carried by a ring in an electron ring accelerator is small compared to the negative charge.] The unknown current in the plates is J_{sy}; we note that it may be represented in the same form [Eq. (9)] as J_x in the previous problem and that its transform also satisfies Eq. (13).

Thus Eq. (A1) differs in structure from Eq. (3) only in having somewhat fewer derivatives, and the Fourier transform procedure of Eqs. (8)–(17) may be carried through with only minor changes. In this way we obtain the analogs of Eq. (14),

$$(k^2+\alpha_n^2)\mathcal{E}_n(k,\omega) = 4\pi i\omega\tilde{j}_- + 4\pi i\omega\frac{qv'}{v}\delta_{n0}e^{ikx_0}, \tag{A3}$$

and of Eq. (19),

$$\frac{E_+}{V_+}(k-i\alpha_0)-\frac{4\pi i\omega qv'e^{ikx_0}}{vV_+(k+i\alpha_0)} = 4\pi i\omega\tilde{j}_-V_-(k-i\alpha_0). \tag{A4}$$

Here,of course,the unknowns E_+ and $\tilde{j}_-$ have different physical meanings from the moving charge case: They refer to y components rather than x components of the field and current. More significantly, the absence in Eq. (A4) of the factor $(k^2-\omega^2)$ on the right-hand side allows us to conclude, by the usual Wiener–Hopf argument, that both sides must equal a constant [rather than, as in the case of Eq. (19), a first-degree polynomial]. The constant may be evaluated as usual by setting $k = i\alpha_0$; thus $\tilde{j}_-$ is determined, and Eq. (A3) yields

$$\mathcal{E}_{sn}(x,\omega) = -\frac{qv'\omega e^{-\alpha_0 x_0}}{v\alpha_0 V_+(i\alpha_0)} \times\int_{-\infty}^{\infty}dk\,\frac{e^{ikx}}{(k-i\alpha_0)V_-(k)(k^2+\alpha_n^2)}, \tag{A5}$$

where the subscript s again indicates that we have omitted the infinite space solution.

Finally we compute the time-averaged energy loss per plate:

$$U' \equiv -\frac{2\pi L}{q^2v}\lim_{T\to\infty}\frac{1}{T}\int_{-T}^{T}dt\int dx\,dz\mathbf{J}_0\cdot\mathcal{E}_s, \tag{A6}$$

where $\mathbf{J}_0$ is,of course,the *given* current and we use a prime to distinguish the energy loss in the moving current case. Equation (A5) yields, in the usual way,

$$U' = -4\pi v'^2\gamma\,\mathrm{Im}\int_0^\infty d\lambda\exp\left[-\frac{2x_0}{L\gamma}\lambda\right]P^2(\lambda,\gamma), \tag{A7}$$

where $P(\lambda,\gamma)$ is defined by Eq. (34).

Comparing the exact Eqs. (36) and (A7), we observe that the relation between U' and U is analytically very simple; in particular, both quantities involve the same integral. It follows that our asymptotic evaluations of U need be only trivially modified.

Specifically in the ultrarelativistic case we find [remember $\gamma = (1-v_z^2)^{-\frac{1}{2}}$]

$$U' = v'^2\gamma^2[a'\gamma^{-\frac{1}{2}}+b'\gamma^{-1}+c'\gamma^{-\frac{3}{2}}+O(\gamma^{-2})], \tag{A8}$$

where

$$\begin{aligned} a' &= a = 0.516\rho^{\frac{3}{2}},\\ b' &= -\rho(1+0.339\rho),\\ c' &= \rho^{\frac{1}{2}}(0.462+0.516\rho+0.0784\rho^2), \end{aligned} \tag{A9}$$

and we recall $\rho = 2\pi L/x_0$. Notice the surprising similarity between b', c' and b, c [cf. Eqs. (65)–(67)].

Note that if we accelerate a rod in the z direction, then the y momentum $(v'\gamma)$ is invariant. Hence if the rod before acceleration ($\gamma = 1$) is relativistic, then $v'\gamma = 1$, and we may expect the radiation due to the y current to equal that from the charge in the limit $\gamma \gg 1$.

In the case of small v, Eq. (A7) implies

$$U' = 8\pi v'^2U_1 = O(v^3), \tag{A10}$$

where U_1 is given by Eq. (98).

APPENDIX B: ENERGY BALANCE ARGUMENTS

In this appendix we present arguments which yield a lower limit to the amount of diffraction radiation produced by a charge Q passing at constant speed v through an accelerating structure of finite dimensions. The discussion is only barely novel; related arguments have been made by Keil, Lawson, and others.

The net energy gain in traversing the structure, ΔU, may, for an electromagnetically linear device, be written in the form

$$\Delta U = AQ - BQ^2. \tag{B1}$$

The coefficient A is proportional to the applied field:

$$A = k_E \mathcal{E}, \tag{B2}$$

where $\mathcal{E}$ is the field, in the absence of the charge Q, measured at some reference position. The coefficient B is the quantity we wish to bound.

The accelerating structure has a total stored energy W, prior to the introduction of the charge Q, which is proportional to $\mathcal{E}^2$:

$$W = k_W \mathcal{E}^2. \tag{B3}$$

Clearly by energy conservation

$$\Delta U \leq W, \tag{B4}$$

which implies, by Eqs. (B1), (B2), and (B3),

$$B \geq \frac{A}{Q^2}\left(Q - \frac{k_W}{k_E^2} A\right). \tag{B5}$$

Taking the maximum, with respect to Q, of the right-hand side of Eq. (B5), yields the limit:

$$B \geq k_E^2/4k_W. \tag{B6}$$

Physically, it is clear that k_W is finite, and it is also clear that there exist accelerating structures for which k_E is nonzero. In particular, even for extreme relativistic particles an efficient acceleration column can be designed; i.e., k_E need *not* decrease with increasing charge speed. For *these structures*—which are just the structures of physical interest—it follows, from Eq. (B6), that B can not decrease without limit with increasing charge speed. The restriction to electromagnetically linear structures is not a severe restriction; one can, for example, imagine disconnecting a structure from the—generally nonlinear—power supplies *after* it has been excited.

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

[1] V. I. Veksler *et al.*, in *Proceedings of the Sixth International Conference on Accelerators, Cambridge, Massachusetts, September 1967*, Cambridge Electron Accelerator Laboratory Report CEAL-2000, p. 289.

[2] For a brief summary of results and references, see J. D. Lawson, Rapporteur's paper, in *Proceedings of the Seventh International Conference on Accelerators, Yerevan, USSR, September 1969* (to be published).

[3] B. S. Levine and A. M. Sessler, IEEE Trans. Nucl. Sci. **16**, 1031 (1969).

[4] O. A. Kolpakov and V. I. Kotov, Zh. Tech. Fiz. **34**, 1387,(1964) [Sov. Phys. Tech. Phys. **9**, 1072 (1965)].

[5] J. D. Lawson, "Radiation from a Ring Charge Passing through a Resonator," Rutherford Laboratory Memorandum RHEL/M144, Rutherford High Energy Laboratory, 1968.

[6] E. D. Courant, Brookhaven National Laboratory, private communication.

[7] G. V. Voskresenskii and B. M. Bolotovskii, Dokl. Akad. Nauk SSSR **154**, 770 (1964) [Sov. Phys. Dokl. **9**, 459 (1964)].

[8] B. M. Bolotovskii and G. V. Voskrenskii, Usp. Fiz. Nauk **94**, 377 (1968) [Sov. Phys. Usp. **11**, 143 (1968)].

[9] A. B. Kuznetsov and S. B. Rubin, Joint Institute for Nuclear Research Report, JINR-P9-4728, Dubna, 1969.

[10] E. Schroeder, Lawrence Radiation Laboratory, Report UCRL-18792, March 1969.

[11] E. Keil, CERN, Internal Report ISR/TH/69-49, 1969, and in *Proceedings of the Seventh International Conference on Accelerators, Yerevan, USSR, September 1969* (to be published).

[12] S. Sackett, Lawrence Radiation Laboratory, private communication.

[13] J. F. Carlson and A. E. Heins, Quart. Appl. Math. **4**, 313 (1947); **5**, 82 (1947).

[14] A. E. Heins, Quart. Appl. Math. **8**, 281 (1950); J. Math. Mech. **6**, 401 (1957).

[15] See, for example, B. Noble, *Methods Based on the Wiener–Hopf Technique* (Pergamon, New York, 1958).

[16] This circumstance was not appreciated in Ref. 7; it was corrected, however, by the same authors in Ref. 8 (see note on p. 151), and independently by the first two authors of this paper, and by Professor D. S. Jones, to whom one of us (A. M. S) is grateful for a private communication.

[17] E. C. Titchmarsh, *The Theory of Functions* (Oxford U.P., Fair Lawn, N.J., 1932).

[18] The representation for $\zeta(1/2)$ appearing in Eq. (59) can be found in: E. C. Titchmarsh, *The Theory of the Riemann Zeta Function* (Oxford U.P., Fair Lawn, N.J., 1951).

[19] See, for example, E. Jahnke and F. Emde, *Tables of Functions* (Dover, New York, 1945).

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 12, NUMBER 3 MARCH 1971

Perturbation Treatment of the Bethe–Salpeter Equation for Tightly Bound Fermion–Antifermion Systems

R. F. KEAM
Department of Physics, University of Auckland, Auckland, New Zealand

(Received 24 June 1970)

A perturbation theory is developed for the ladder approximation Bethe–Salpeter equation that describes bound spin-$\frac{1}{2}$ fermion–antifermion systems of small total mass. A new exact analytical solution for a zero-mass system is presented and the perturbation theory is applied to it. Some time-reversal properties of the equation are also examined.

I. INTRODUCTION

In the following sections we develop a perturbation treatment of the Bethe–Salpeter equation[1,2] for a bound spin-$\frac{1}{2}$ fermion–antifermion system. Starting with any exact analytical solution which satisfies certain boundary conditions and describes the bound system for zero total mass, we find formulas for the approximate analytic dependence of the coupling constant on the bound state mass and on the difference in mass of the bound particles for small values of these two parameters. We have not attempted to explore the effects of allowing the potential to vary.

Ito[3] has also recently examined perturbation of the zero-total-mass equation (for the equal mass case only), but the lowest-order nonvanishing correction term for the coupling constant must include contributions from the perturbed Bethe–Salpeter amplitude,[4] and these unfortunately have been overlooked in his treatment.

After recapitulating certain properties of the Bethe–Salpeter equation in Sec. II, in Sec. III we examine its time-reversal properties. The perturbation equations are set up in Sec. IV, and developed in detail for S–V sector solutions in Sec. V. Finally, in Sec. VI a new exact solution of the unperturbed equation is presented, and as an example the theory of the previous sections is applied to it to obtain the relevant perturbation coefficients.

The notation and conventions are the same as in previous papers[5–7] on similar topics by the author. We denote complex conjugate by *, Hermitian conjugate by †, and transpose by ^{T}.

II. BETHE–SALPETER EQUATION

We consider the bound system of total 4-momentum P formed from a spin-$\frac{1}{2}$ fermion of mass m_a and a spin-$\frac{1}{2}$ antifermion of mass m_b. If the state vector for this system is $|B\rangle$, the corresponding 2-body amplitude may be written as

$$\begin{aligned}\chi_{\alpha\beta}(x_1, x_2) &= \langle 0|\, T\{\psi_\alpha(x_1)\bar{\psi}_\beta(x_2)\}\,|B\rangle \\ &= e^{iP\cdot X}\langle 0|\, T\{\psi_\alpha(\mu_b x)\bar{\psi}_\beta(-\mu_a x)\}\,|B\rangle \\ &= e^{iP\cdot X} f_{\alpha\beta}(x), \end{aligned} \tag{1}$$

where

$$\begin{aligned} X &= \mu_a x_1 + \mu_b x_2, \\ x &= x_1 - x_2 \end{aligned} \tag{2}$$

for arbitrary real values of μ_a, μ_b satisfying

$$\mu_a + \mu_b = 1. \tag{3}$$

The adjoint amplitude is

$$\begin{aligned}\bar{\chi}_{\beta\alpha}(x_2, x_1) &= \langle B|\, T\{\bar{\psi}_\alpha(x_1)\psi_\beta(x_2)\}\,|0\rangle \\ &= e^{-iP\cdot X}\langle B|\, T\{\bar{\psi}_\alpha(\mu_b x)\psi_\beta(-\mu_a x)\}\,|0\rangle \\ &= e^{-iP\cdot X}\bar{f}_{\beta\alpha}(x). \end{aligned} \tag{4}$$

If

$$\begin{aligned} f(p) &= \int d^4x e^{-ip\cdot x} f(x), \\ \bar{f}(p) &= \int d^4x e^{ip\cdot x}\bar{f}(x), \end{aligned} \tag{5}$$

then

$$\bar{f}(\mathbf{p}, p^{0*}) = \gamma_4[f(\mathbf{p}, p^0)]^\dagger\gamma_4 \tag{6}$$

except for p^0 on the real axis.

For the bound system we write the Wick-rotated, ladder-approximation, relative configuration space Bethe–Salpeter equation as

$$\begin{aligned}\{\gamma\cdot(-i\partial + \mu_a P) - im_a\}f(x)\{\gamma\cdot(-i\overleftarrow{\partial} - \mu_b P) - im_b\} \\ = \lambda\mathcal{U}f(x) \end{aligned} \tag{7}$$

or, briefly,

$$\mathcal{B}_P f(x) = -\lambda\mathcal{U}f(x). \tag{8}$$

In the rest system these become

$$\begin{aligned}(\gamma\cdot\partial - \mu_a E + m_a)f(x)(\gamma\cdot\overleftarrow{\partial} + \mu_b E + m_b) \\ = -\lambda\mathcal{U}f(x), \end{aligned}$$

$$\mathcal{B}_E f(x) = -\lambda\mathcal{U}f(x). \tag{9}$$

The linear operator $\mathcal{U}$, which describes the interaction binding the particles, we assume to be radial, i.e., to be of form [cf. Eq. (I.12)]

$$\mathcal{U}f = \sum_i V_i(R)\hat{\Gamma}_i f, \tag{10}$$

where R is the 4-dimensional distance from the origin of the Euclidean relative configuration space.

Alternatively, in the Euclidean relative momentum space we have

$$\{\gamma\cdot(p+\mu_a P)-im_a\}f(p)\{\gamma\cdot(p-\mu_b P)-im_b\} = \lambda\mathfrak{U}(f,p) \quad (11)$$

or

$$\mathcal{A}_P f(p) = -\lambda\mathfrak{U}(f,p). \quad (12)$$

$\mathcal{A}_P f(p)$ is the Fourier transform of $\mathcal{B}_P f(x)$ and $\mathfrak{U}$ is a linear integral operator:

$$\mathfrak{U}(f,p) = (2\pi)^{-4}\int d^4k\,\mathfrak{W}(|p-k|)f(k), \quad (13)$$

where $\mathfrak{W}$ is the Fourier transform of $\mathfrak{V}(R)$. Thus,

$$\mathfrak{W}(K) = \int d^4x e^{-ik\cdot x}\mathfrak{V}(R) = (2\pi)^2K^{-1}\int_0^\infty dR R^2 J_1(KR)\mathfrak{V}(R), \quad (14)$$

and we emphasize the fact that it is a function of the radial coordinate K only, in momentum space. For instance, if the potential is due to exchange of a boson of mass μ, and type i, $\mathfrak{W}$ is proportional to the propagator, and indeed

$$\mathfrak{W}(K) = -\frac{(4\pi)^2}{K^2+\mu^2}\epsilon\hat{\Gamma}_i, \quad (15)$$

where ϵ is given in Table II of Paper II.

If we derive the Bethe–Salpeter equation for the adjoint $\bar{f}(p)$, we obtain the Wick-rotated form

$$\{\gamma\cdot(p-\mu_b P)-im_b\}\bar{f}(p)\{\gamma\cdot(p+\mu_a P)-im_a\} = \lambda\mathfrak{U}(\bar{f},p). \quad (16)$$

Alternatively, if we take the Hermitian conjugate of Eq. (8), pre- and postmultiply by γ_4, and use

$$\bar{f}(\mathbf{p},-p_4) = \gamma_4 f^\dagger(\mathbf{p},p_4)\gamma_4, \quad (17)$$

which is derivable from (6), we obtain (16) with $\lambda^*\mathfrak{U}^\dagger$ replacing the right member. Provided each $V_i(R)$ in (10) is real, $\mathfrak{U}^\dagger = \mathfrak{U}$. We see that if λ is unique, it is real. The result that λ is real in the general case has been obtained by Naito and Nakanishi.[8]

III. TIME REVERSAL PROPERTIES

We now wish to identify the operator $\overline{\mathcal{C}}$ which, acting on the state amplitude χ, is equivalent to the Hilbert space time-reversal operator $\bar{T}$ acting on the state vector $|B\rangle$, i.e.,

$$\overline{\mathcal{C}}\chi = \langle 0|\,T\{\psi(x_1)\bar{\psi}(x_2)\}\bar{T}\,|B\rangle. \quad (18)$$

(In this, T is the Wick chronological product operator which orders ψ, $\bar{\psi}$, and is not related to $\bar{T}$.)

We assume the relations

$$\bar{T}^{-1}\psi_\alpha(x)\bar{T} = B\psi^*_\alpha(x'),$$
$$\bar{T}^{-1}\bar{\psi}_\beta(x)\bar{T} = \bar{\psi}^*_\beta(x')B^{-1}, \quad (19)$$

where the 4×4 matrix B has the properties

$$\gamma_\mu^{\mathrm{T}} = B^{-1}\gamma_\mu B,$$
$$B^{\mathrm{T}} = -B, \quad B^\dagger = B^{-1}. \quad (20)$$

In (19), if $x=(\mathbf{x},t)$, then $x'=(\mathbf{x},-t)$.

Using (1), (4), and (18)–(20), we deduce that

$$\overline{\mathcal{C}}\chi(x_1,x_2) = -\gamma_4 B\bar{\chi}^{\mathrm{T}}(x_2',x_1')B^{-1}\gamma_4,$$
$$\overline{\mathcal{C}}\bar{\chi}(x_2,x_1) = -\gamma_4 B\chi^{\mathrm{T}}(x_1',x_2')B^{-1}\gamma_4, \quad (21)$$

or, in terms of the wavefunction in relative configuration space,

$$\overline{\mathcal{C}}f(x) = -\gamma_4 B\bar{f}^{\mathrm{T}}(x')B^{-1}\gamma_4,$$
$$\overline{\mathcal{C}}\bar{f}(x) = -\gamma_4 B f^{\mathrm{T}}(x')B^{-1}\gamma_4. \quad (22)$$

In relative momentum space these become

$$\overline{\mathcal{C}}f(p) = -\gamma_4 B\bar{f}^{\mathrm{T}}(-p')B^{-1}\gamma_4,$$
$$\overline{\mathcal{C}}\bar{f}(p) = -\gamma_4 B f^{\mathrm{T}}(-p')B^{-1}\gamma_4, \quad (23)$$

where $p'=(-\mathbf{p},p^0)$, the time-reversed p.

Since for complex p^0 we have Eq. (6), we can deduce for real p_4

$$\overline{\mathcal{C}}f(\mathbf{p},p_4) = -Bf^*(\mathbf{p},p_4)B^{-1}. \quad (24)$$

Again, taking p_4 real and the Fourier transform in the 4-dimensional Euclidean space, we see that (6) implies

$$\bar{f}(\mathbf{x},-x_4) = \gamma_4 f^\dagger(\mathbf{x},x_4)\gamma_4 \quad (25)$$

and hence from (22)

$$\overline{\mathcal{C}}f(\mathbf{x},x_4) = -Bf^*(\mathbf{x},x_4)B^{-1}. \quad (26)$$

From the occurrence of the complex conjugate in (24) and (26), it is clear that $\overline{\mathcal{C}}$ is an antilinear operator.

We now examine the time-reversal properties of (7), when $\mathfrak{V}$ is given by (10). According to the rules[9] we obtain the time-reversed equation by taking the complex conjugate of (7), noting that the time-reversed total momentum $P'=(-\mathbf{P},iP^0)=-P^*$ and that $\partial'=\partial^*=\partial$.

Equation (7) becomes

$$\{\gamma^*\cdot(i\partial'-\mu_a P')+im_a\}f^*(x) \times\{\gamma^*\cdot(i\overleftarrow{\partial'}+\mu_b P')+im_b\} = \lambda^*\mathfrak{V}^*f^*(x). \quad (27)$$

Using (20) and (26) with $f'(x)=\overline{\mathcal{C}}f(x)$, we find

$$\{\gamma\cdot(-i\partial'+\mu_a P')-im_a\}f'(x) \times\{\gamma\cdot(-i\partial'-\mu_b P')-im_b\} = \lambda^*\sum_i V_i^*(R)\Gamma_i f'(x). \quad (28)$$

Comparing (28) with (7), we see that the latter is time-reversal invariant provided that $\lambda^* V_i^* = \lambda V_i$ for each i, a condition we have already found to hold in Sec. II.

Following the notation of Paper I, we note that the linear operators $\mathbf{L}^+$, $\mathbf{S}^+$, $\mathbf{L}^-$, $\mathbf{S}^-$, $\mathbf{J}^+ = \mathbf{L}^+ + \mathbf{S}^+$, $\mathbf{J}^- = \mathbf{L}^- + \mathbf{S}^-$, and $\mathbf{J} = \mathbf{J}^+ + \mathbf{J}^-$ have the properties of a set of commuting angular momenta. Thus adopting standard notation[10] for the generalized angular momenta, we may write $Z_{lm^+m^-} = |lm^+\rangle\,|lm^-\rangle = |llm^+m^-\rangle$, where $Z_{lm^+m^-}$ is defined in Eq. (I.35). The properly phased Dirac-space states obtained from Table II of I, when normalized, we may similarly denote as $|\Gamma_i\, s^+\, s^-\, m_s^+\, m_s^-\rangle$, where Γ_i identifies the particular sector of Dirac space and so removes a possible ambiguity. The angular and Dirac-space coefficients occurring (e.g.) in $f_{j+j-m+m-}$ of Eq. (I.38) may be written $|\Gamma_i\, l^+\, s^+, j^+\, m^+\rangle\,|\Gamma_i\, l^-\, s^-, j^-\, m^-\rangle$ or $|\Gamma_i(l^+\, s^+)j^+(l^-\, s^-)j^-; m^+\, m^-\rangle$, and finally the angular and Dirac-space coefficients occurring in f_{j+j-Jm} of Eq. (I.54) may be written $|\Gamma_i(l^+\, s^+)j^+(l^-\, s^-)j^-; J\, m\rangle$. (For brevity we may sometimes omit the $l^\pm$, $s^\pm$, where all four are coupled, and put simply $|\Gamma_i\, j^+\, j^-\, m^+\, m^-\rangle$ or $|\Gamma_i\, j^+\, j^-\, J\, m\rangle$ as the case may be.)

We define the bra vector corresponding to any angular and/or Dirac-space ket vector (in the sense used in this paragraph) to be the Hermitian conjugate of that ket:

$$\langle A| = |A\rangle^\dagger. \tag{29}$$

The scalar product is defined as the operation of taking the trace in Dirac space and integrating over all angles in configuration space. By (A7) of Appendix A we thus have

$$\langle l\, l\, m^+\, m^- \,|\, k\, k\, n^+\, n^-\rangle = \delta_{lk}\delta_{m^+n^+}\delta_{m^-n^-}. \tag{30}$$

Similarly

$$\langle \Gamma_i\, s^+\, s^-\, m^+\, m^- \,|\, \Gamma_j\, r^+\, r^-\, n^+\, n^-\rangle = \delta_{ij}\delta_{s^+r^+}\delta_{s^-r^-}\delta_{m^+n^+}\delta_{m^-n^-}. \tag{31}$$

The coupled angular-momentum vectors defined above are as a consequence also clearly orthonormal in their respective arguments.

The following commutation and anticommutation properties of $\overline{\mathcal{C}}$ are readily verified:

$$[\overline{\mathcal{C}}, (\mathbf{L}^\pm)^2] = [\overline{\mathcal{C}}, (\mathbf{S}^\pm)^2] = [\overline{\mathcal{C}}, (\mathbf{J}^\pm)^2] = [\overline{\mathcal{C}}, \mathbf{J}^2] = 0,$$
$$[\overline{\mathcal{C}}, L_i^\pm]_+ = [\overline{\mathcal{C}}, S_i^\pm]_+ = [\overline{\mathcal{C}}, J_i^\pm]_+ = [\overline{\mathcal{C}}, J_i]_+ = 0. \tag{32}$$

Thus applied to a simultaneous eigenstate of any commuting set of the generalized and/or ordinary angular momenta and their z components, $\overline{\mathcal{C}}$ produces an eigenstate with the same total angular momenta but with reversed components. From the unitary character of $\overline{\mathcal{C}}$ the normalization is unaltered, but a phase change may have been introduced. Indeed if $Z(j_i, m_j)$ is an arbitrary angular variable and/or Dirac-space angular momentum eigenfunction corresponding to quantum numbers j_i and m_j, then

$$\overline{\mathcal{C}}Z(j_i, m_j) = (-1)^\eta Z(j_i, -m_j). \tag{33}$$

In Table I we list the value of η for each commonly occurring form of $Z(j_i, m_j)$.[5,11]

Since $\overline{\mathcal{C}}$ is antilinear, the reality of the Clebsch-Gordan coefficients ensures that η is well defined.

As in Paper I, zero-mass solutions of the Bethe-Salpeter equation for the system we are considering are of the form

$$f^\Gamma_{j^+j^-m^+m^-} = \sum_l f_l(R)\,|\Gamma_i(ls^+)j^+(ls^-)j^-; m^+\, m^-\rangle \tag{34}$$

or

$$f^\Gamma_{j^+j^-Jm} = \sum_l f_l(R)|\Gamma_i(ls^+)j^+(ls^-)j^-; J\, m\rangle. \tag{35}$$

Since the value of η in each sum can easily be shown to be independent of l, we have in each case

$$\overline{\mathcal{C}}f^\Gamma_{j^+j^-m^+m^-} = (-1)^\eta \sum f_l^*(R)\,|\Gamma_i(ls^+)j^+(ls^-)j^-; -m^+\, -m^-\rangle. \tag{36}$$

But

$$f^\Gamma_{j^+j^--m^+-m^-} = \sum_l f_l(R)\,|\Gamma_i(ls^+)j^+(ls^-)j^-; -m^+\, -m^-\rangle. \tag{37}$$

On physical grounds we clearly wish to identify the two states $\overline{\mathcal{C}}f^\Gamma_{j^+j^-m^+m^-}$ and $f^\Gamma_{j^+j^--m^+-m^-}$, and this is

TABLE I. Phase factors $(-1)^\eta$ introduced into the bispherical angular and Dirac space eigenfunctions by the time-reversal operator $\overline{\mathcal{C}}$ according to Eq. (33), $m = m^+ + m^-$.

$Z(j_i, m_j)$		η
$\|l\, l\, m_l^+\, m_l^-\rangle$		$m_l^+ + m_l^- + 1$
$\|\Gamma_i\, s^+\, s^-\, m_s^+\, m_s^-\rangle$	$i \neq 4$	$m_s^+ + m_s^- + 1$
$\|\Gamma_4\, s^+\, s^-\, m_s^+\, m_s^-\rangle$		$m_s^+ + m_s^-$
$\|\Gamma_i\, j\, j\, m^+\, m^-\rangle$	$i = 1, 2, 5$	$m + 1$
	$i = 3, 4$	m
$\|\Gamma_i\, j\, j + 1\, m^+\, m^-\rangle$	$i = 2$	m
	$i = 3, 4$	$m + 1$
$\|\Gamma_i\, j + 1\, j\, m^+\, m^-\rangle$	$i = 2$	m
	$i = 3, 4$	$m + 1$
$\|\Gamma_i\, j\, j\, J\, m\rangle$	$i = 1, 2, 5$	$2j - J + m + 1$
	$i = 3, 4$	$2j - J + m$
$\|\Gamma_i\, j\, j + 1\, J\, m\rangle$	$i = 2, 3, 4$	$2j - J + m$
$\|\Gamma_i\, j + 1\, j\, J\, m\rangle$	$i = 2, 3, 4$	$2j - J + m$

achieved by putting

$$f_l^*(R) = e^{i\zeta} f_l(R), \tag{38}$$

where ζ is independent of l and may as well be chosen zero. Thus we may choose all configuration space radial functions to be real. Such a choice is clear enough from Eqs. (I.43)–(I.53), but we can now attribute it to the time-reversal properties of the interaction.

In relative-momentum space on the other hand, because the Hankel transform equations [(III.11a) and (III.11b)] introduce factors $i^{\pm 2l}$, the momentum-space radial functions separate into two disjoint sets $\{f_S, f_T, f_P\}$ and $\{f_V, f_A\}$, any member in one of which is real relative to another member of the same set and purely imaginary relative to a member of the other set.

IV. PERTURBATION EQUATIONS

As was shown in Paper I, solutions of the Bethe–Salpeter equation for zero total mass may be chosen in the case $m_a = m_b$ to be simultaneous eigenstates of the operator sets $\{\alpha, \beta, \mathbf{J}^2, J_z, \mathcal{R}\}$ or equivalently $\{(\mathbf{J}^+)^2, (\mathbf{J}^-)^2, \mathbf{J}^2, J_z, \mathcal{R}\}$. Either set with the relevant Bethe–Salpeter operator $\mathcal{B}_0^{-1}\mathcal{U}$ adjoined constitutes a complete set of commuting operators for the system. Simultaneous eigenfunctions of these complete sets are, in general, not parity eigenfunctions since one can show [e.g., from (II.8)]

$$[\bar{\mathcal{T}}, \beta]_+ = 0. \tag{39}$$

One has immediately, however, that $\bar{\mathcal{T}}$ commutes with β^2, so that a complete set including $\bar{\mathcal{T}}$ is given by

$$\{\mathcal{B}_0^{-1}\mathcal{U}, \alpha, \beta^2, \mathbf{J}^2, J_z, \bar{\mathcal{T}}, \mathcal{R}\}. \tag{40}$$

Simultaneous eigenfunctions of each of these sets have a well-defined charge parity since for zero-mass solutions we have the relation (II.7), and the actual forms of the eigenfunctions follow from Sec. 2 and Table I of Paper II and from Appendix E. The eigenvalue of $\mathcal{B}_0^{-1}\mathcal{U}$ is $-\lambda^{-1}$, where λ is the coupling constant. Let us distinguish between different possible eigenvalues for a given set of eigenvalues of the members of the set other than $\mathcal{B}_0^{-1}\mathcal{U}$ by an index i. As is apparent, e.g., from the radial equations (I.43)–(I.53), λ is independent of the eigenvalues of $\mathbf{J}^2$ and J_z. Thus

$$\lambda = \lambda(i, \alpha, \beta^2, \bar{\mathcal{T}}, \mathcal{R}, m), \tag{41}$$

where we have made explicit the dependence through $\mathcal{B}_0^{-1}\mathcal{U}$ of λ on the (equal) masses m of the interacting fermion and antifermion. Notice that, since $\mathcal{B}_0$ is of dimension (mass)^2, if $\mathcal{U}$ is of dimension (mass)$^{2-r}$,

$$\lambda(i, \alpha, \beta^2, \bar{\mathcal{T}}, \mathcal{R}, m) = m^r \bar{\lambda}(i, \alpha, \beta^2, \bar{\mathcal{T}}, \mathcal{R}), \tag{42}$$

where $\bar{\lambda}$ is independent of m provided any other masses in $\mathcal{U}$ are changed in proportion with m. When $\mathcal{U}$ is the potential due to exchange of a particle, $r = 0$ and $\bar{\lambda} \equiv \lambda$. We shall assume this to be the case hereafter.

When the total mass E is nonzero and/or $m_a \neq m_b$, $\mathcal{B}_E$ no longer commutes with α or β. It still commutes with $\mathbf{J}^2$, J_z, $\bar{\mathcal{T}}$, and $\bar{\mathcal{C}}$ the charge parity operator. We shall assume that, given any simultaneous eigenfunction of the operators in set (40), we can perturb this to find an approximate solution when E is small but nonzero and when $m_a \sim m_b$. Such a solution we shall require to retain the values of $\mathbf{J}^2$, J_z, $\bar{\mathcal{T}}$, and $\bar{\mathcal{C}}$ that the unperturbed solution had.

Putting

$$\begin{gathered} \mu_a = u_b = \tfrac{1}{2}, \\ m_a = m + \Delta, \quad m_b = m - \Delta, \\ \epsilon = \tfrac{1}{2}E, \end{gathered} \tag{43}$$

we see that the form of Eq. (11) in the center-of-mass system becomes

$$\begin{aligned} &(\gamma \cdot p + i\epsilon\gamma_4 - im - i\Delta) f(p) \\ &\quad \times (\gamma \cdot p - i\epsilon\gamma_4 - im + i\Delta) = \lambda \mathcal{U}(f, p) \end{aligned} \tag{44}$$

or briefly

$$\mathcal{A}_{\epsilon,\Delta} f(p) = -\lambda \mathcal{U}(f, p). \tag{45}$$

λ now clearly is a function of $\bar{\mathcal{T}}$, $\bar{\mathcal{C}}$, ϵ, and Δ, which in the limit $\epsilon \to 0$, $\Delta \to 0$ becomes $\lambda(i, \alpha, \beta^2, \bar{\mathcal{T}}, \mathcal{R})$. We shall write only the variables ϵ and Δ explicitly, viz.,

$$\lambda = \lambda(\epsilon, \Delta). \tag{46}$$

Let us assume that λ is an analytic function of ϵ and Δ about $\epsilon = \Delta = 0$. Then we have

$$\lambda = \sum_{r,s=0}^{\infty} \lambda_{rs} \epsilon^r \Delta^s, \tag{47}$$

where

$$\lambda_{rs} = \frac{1}{r!\, s!} \frac{\partial^r}{\partial \epsilon^r} \frac{\partial^s}{\partial \Delta^s} \lambda \bigg|_{\epsilon=\Delta=0} \tag{48}$$

The corresponding momentum-space eigenfunction $f(p)$ also depends on ϵ and Δ, and, assuming analyticity in these variables about their zero values, we can write

$$f = \sum_{r,s=0}^{\infty} f_{rs} \epsilon^r \Delta^s. \tag{49}$$

Substituting the power series (47) and (49) in (44), we obtain by equating coefficients of like powers of

ϵ and Δ the following set of equations:

$$(\mathcal{A}_{00} + \lambda_{00}\mathcal{U})f_{00} = 0, \tag{50_{00}}$$

$$(\mathcal{A}_{00} + \lambda_{00}\mathcal{U})f_{10} = i\{\gamma_4 f_{00}(\gamma\cdot p - im) - (\gamma\cdot p - im)f_{00}\gamma_4\} - \lambda_{10}\mathcal{U}f_{00}, \tag{50_{10}}$$

$$(\mathcal{A}_{00} + \lambda_{00}\mathcal{U})f_{01} = -i\{f_{00}\gamma\cdot p - \gamma\cdot p f_{00}\} - \lambda_{01}\mathcal{U}f_{00}, \tag{50_{01}}$$

and, in general,

$$\begin{aligned}(\mathcal{A}_{00} + \lambda_{00}\mathcal{U})f_{rs} &= i\{\gamma_4 f_{r-1\,s}(\gamma\cdot p - im) - (\gamma\cdot p - im)f_{r-1\,s}\gamma_4\} \\ &\quad - i\{f_{r\,s-1}\gamma\cdot p - \gamma\cdot p f_{r\,s-1}\} \\ &\quad + \gamma_4 f_{r-2\,s}\gamma_4 + f_{r\,s-2} - \{\gamma_4 f_{r-1\,s-1} + f_{r-1\,s-1}\gamma_4\} \\ &\quad - \sum_{\substack{m=0\\ m+n\geq 1}}^{r}\sum_{n=0}^{s}\lambda_{mn}\mathcal{U}f_{r-m\,s-n},\end{aligned} \tag{50_{rs}}$$

where in the right member any f_{ab} with negative a or b is put equal to zero. In principle, the successive solution of these equations would allow us to determine f to any required order of ϵ and Δ.

Of more interest, however, is the expansion (47) for λ.

Define the linear function $\mathfrak{L}\{F, \}$, acting on 4×4 matrix functions $G(p)$ in Euclidean relative momentum space, by

$$\mathfrak{L}\{F, G\} = \int d^4p \,\mathrm{Tr}\,[F(p)G(p)], \tag{51}$$

where the integral is over the complete space and where one assumes that F and G have the properties necessary for the integral to converge. F also is to be a 4×4 matrix function. For arbitrary F(resp. G), we define $\bar{F}$(resp. $\bar{G}$), by Eq. (17).

Now

$$\mathfrak{L}\{\bar{f}_{00}, (\mathcal{A}_{00} + \lambda_{00}\mathcal{U})f_{rs}\} = \mathfrak{L}\{f_{rs}, (\mathcal{A}_{00} + \lambda_{00}\mathcal{U})\bar{f}_{00}\} = 0 \tag{52}$$

by Eqs. (50_{00}), (16), and (17).

Next consider $\mathfrak{L}\{\bar{f}_{00}, \gamma_4 f_{00}(\gamma\cdot p - im)\}$. It is not difficult to show that the presence of the γ_4 factor causes terms with nonvanishing traces in $\mathfrak{L}$ always to have bispherical angular functions belonging to integral l values multiplied by angular functions belonging to half-odd-integral l values. When the angular integrations are performed, such products must vanish by (A5) of Appendix A. Thus

$$\mathfrak{L}\{\bar{f}_{00}, \gamma_4 f_{00}(\gamma\cdot p - im)\} = 0 \tag{53a}$$

and similarly

$$\mathfrak{L}\{\bar{f}_{00}, (\gamma\cdot p - im)f_{00}\gamma_4\} = 0. \tag{53b}$$

Applying $\mathfrak{L}\{\bar{f}_{00}, \}$ to Eq. (50_{10}) and using (52) and (53), we thus see that, unless $L_{00} \equiv \mathfrak{L}\{\bar{f}_{00}, \mathcal{U}f_{00}\}$ vanishes,

$$\lambda_{10} = 0. \tag{54}$$

Whichever sector (S–V, T–A, P)f_{00} is in, it is easily shown that $f_{00}\gamma\cdot p - \gamma\cdot p f_{00}$ is in a sector or combination of the two sectors orthogonal to it. Therefore from the application of the trace operation

$$\mathfrak{L}\{\bar{f}_{00}, f_{00}\gamma\cdot p - \gamma\cdot p f_{00}\} = 0 \tag{55}$$

and, from (50_{01}) (again unless L_{00} vanishes),

$$\lambda_{01} = 0. \tag{56}$$

From Eq. (50_{11}) we find

$$\begin{aligned}\lambda_{11}L_{00} &= i\mathfrak{L}\{\bar{f}_{00}, \gamma_4 f_{01}(\gamma\cdot p - im) - (\gamma\cdot p - im)f_{01}\gamma_4\} \\ &\quad - i\mathfrak{L}\{\bar{f}_{00}, f_{10}\gamma\cdot p - \gamma\cdot p f_{10}\} \\ &\quad - \mathfrak{L}\{\bar{f}_{00}, \gamma_4 f_{00} + f_{00}\gamma_4\}.\end{aligned} \tag{57}$$

The last contribution on the right vanishes by the same argument from which we obtained (53a). Using the analogs of (50_{10}) and (50_{01}) for $\bar{f}$, we then find

$$\begin{aligned}\lambda_{11}L_{00} &= \mathfrak{L}\{\bar{f}_{10}, (\mathcal{A}_{00} + \lambda_{00}\mathcal{U})f_{01}\} \\ &\quad + \mathfrak{L}\{f_{10}, (\mathcal{A}_{00} + \lambda_{00}\mathcal{U})\bar{f}_{01}\} \\ &= 2\,\mathrm{Re}\,\mathfrak{L}\{\bar{f}_{10}, (\mathcal{A}_{00} + \lambda_{00}\mathcal{U})f_{01}\},\end{aligned} \tag{58}$$

and this vanishes since it can be shown that

$$\mathfrak{L}\{\bar{f}_{10}, (\mathcal{A}_{00} + \lambda_{00}\mathcal{U})f_{01}\}$$

is purely imaginary. Again provided L_{00} is nonzero,

$$\lambda_{11} = 0. \tag{59}$$

We note also from (50_{00}) that

$$L_{00} = -\lambda_{00}^{-1}\mathfrak{L}\{\bar{f}_{00}, \mathcal{A}_{00}f_{00}\}. \tag{60}$$

Thus, for small ϵ and Δ, an expansion for λ to the lowest nonzero contributing orders is

$$\lambda \approx \lambda_{00} + \epsilon^2\lambda_{20} + \Delta^2\lambda_{02}, \tag{61}$$

where, from (50_{20}) and (50_{02}),

$$\begin{aligned}\lambda_{20}L_{00} &= i\mathfrak{L}\{\bar{f}_{00}, \gamma_4 f_{10}(\gamma\cdot p - im) - (\gamma\cdot p - im)f_{10}\gamma_4\} \\ &\quad + \mathfrak{L}\{\bar{f}_{00}, \gamma_4 f_{00}\gamma_4\},\end{aligned} \tag{62}$$

$$\lambda_{02}L_{00} = -i\mathfrak{L}\{\bar{f}_{00}, f_{01}\gamma\cdot p - \gamma\cdot p f_{01}\} + \mathfrak{L}\{\bar{f}_{00}, f_{00}\}. \tag{63}$$

λ_{20} and λ_{02} can thus be found provided that Eqs. (50_{10}) and (50_{01}) can be solved for f_{10} and f_{01} for the particular potential.

A further useful relation can be found in the particular case where the potential is due to the exchange of a massless particle. In this case we do not have to change $\mathcal{U}$ if we change m.[12] We exploit the relation

$\partial\lambda/\partial m = 0$: If one applies $\mathfrak{L}\{\bar{f}_{00}, \ \}$ to the first-order terms of the equation obtained by perturbing the particle and antiparticle masses equally in (50_{00}), one obtains

$$\mathfrak{L}\{\bar{f}_{00}, f_{00}(\gamma\cdot p - im) + (\gamma\cdot p - im)f_{00}\} = 0, \quad (64)$$

which, combined with (55), gives

$$\mathfrak{L}\{\bar{f}_{00}, f_{00}\gamma\cdot p\} = \mathfrak{L}\{\bar{f}_{00}, \gamma\cdot p f_{00}\} = im\mathfrak{L}\{\bar{f}_{00}, f_{00}\}. \quad (65)$$

We note that condition (III.16) is in general necessary for the radial functions in f_{00} in order that $\mathfrak{L}\{\bar{f}_{00}, \mathcal{A}_{00}f_{00}\}$ converge, and hence from (60) in order that L_{00} be finite.

V. APPLICATION TO S–V SECTOR SOLUTIONS

We now restrict our attention to the perturbation of zero-total-mass solutions from the S–V sector of Dirac space. There are three distinct types of such solutions.[13] The corresponding eigenvalues of the operators in set (40), except for the potential-dependent $\mathcal{B}_0^{-1}\mathcal{V}$, are listed in Table II.

In the following analysis we shall consider only f^{SV} type solutions.

From (I.38), (I.54), and Eq. (35) of this paper we have, for $j \neq 0$,

$$\begin{aligned} f_{00}(x) = {} & f_S(R)\,|\Gamma_1(j\,0)j(j\,0)j; J\,m\rangle \\ & + f_{V1}(R)\,|\Gamma_2(j+\tfrac{1}{2}\,\tfrac{1}{2})j(j+\tfrac{1}{2}\,\tfrac{1}{2})j; J\,m\rangle \\ & + f_{V2}(R)\,|\Gamma_2(j-\tfrac{1}{2}\,\tfrac{1}{2})j(j-\tfrac{1}{2}\,\tfrac{1}{2})j; J\,m\rangle \quad (66) \end{aligned}$$

or briefly

$$f_{00}(x) = f_S(R)\,|S\rangle + f_{V1}(R)\,|V1\rangle + f_{V2}(R)\,|V2\rangle. \quad (67)$$

In relative momentum space (cf. the Appendix of Paper III) we similarly have, using k as our momentum variable,

$$f_{00}(k) = f_S(K)\,|S\rangle + f_{V1}(K)\,|V1\rangle + f_{V2}(K)\,|V2\rangle, \quad (68)$$

where, as in (III.11a) and (III.11b), $Kf_l(K)$ and $(2\pi)^2Rf_l(R)$ are mutual Hankel transforms with respect to the Bessel function of order $2l+1$ and where the angular variables in the kets now refer to momentum space.

From (17),

$$\begin{aligned} \bar{f}_{00}(k) &= [\gamma_4 f_{00}(\mathbf{k}, -k_4)\gamma_4]^\dagger \\ &= (-1)^{2j-J}\{\langle S|\, f_S^*(K) \\ &\quad - \langle V1|\, f_{V1}^*(K) - \langle V2|\, f_{V2}^*(K)\} \quad (69a) \\ &= (-1)^J\{\langle S|\, f_S + \langle V1|\, f_{V1} + \langle V2|\, f_{V2}\}, \quad (69b) \end{aligned}$$

where in (69b) we have exploited the fact that configuration space radial functions are real and consequently, through (III.11b), each momentum space radial function has the reality property of $(-i)^{2l}$ for its appropriate l.

From Eqs. (60), (62), and (63) we see that $\mathfrak{L}\{\bar{f}_{00}, f_{00}\}$, $\mathfrak{L}\{\bar{f}_{00}, \gamma_4 f_{00}\gamma_4\}$, and $\mathfrak{L}\{\bar{f}_{00}, \mathcal{A}_{00}f_{00}\}$ must be evaluated.

Using the definition of the scalar product of angular and of Dirac space functions, we have

$$\begin{aligned} \mathfrak{L}\{\bar{f}_{00}, f_{00}\} &= \int d^4k\ \mathrm{Tr}\,[\bar{f}_{00}(k)f_{00}(k)] \\ &= (-1)^{2j-J}\int dK K^3(|f_S|^2 - |f_{V1}|^2 - |f_{V2}|^2). \end{aligned} \quad (70)$$

From the analysis of Appendix C and the detailed calculations in Appendix D,

$$\begin{aligned} &(-1)^{2j-J}\mathfrak{L}\{\bar{f}_{00}, \gamma_4 f_{00}\gamma_4\} \\ &\quad = \int dK K^3[|f_S|^2\langle S|\,(\gamma_4\times\gamma_4)\,|S\rangle \\ &\qquad - |f_{V1}|^2\langle V1|\,(\gamma_4\times\gamma_4)\,|V1\rangle \\ &\qquad - |f_{V2}|^2\langle V2|\,(\gamma_4\times\gamma_4)\,|V2\rangle] \\ &\quad = \int dK K^3\left[|f_S|^2 + |f_{V1}|^2\left(\frac{2(j+1)+J(J+1)}{4(j+1)^2}\right)\right. \\ &\qquad \left. - |f_{V2}|^2\left(\frac{2j-J(J+1)}{4j^2}\right)\right]. \end{aligned} \quad (71)$$

Also

$$\begin{aligned} \mathfrak{L}\{\bar{f}_{00}, -\mathcal{A}_{00}f_{00}\} &= \mathfrak{L}\{\bar{f}_{00}, \gamma\cdot p f_{00}\gamma\cdot p\} \\ &\quad - im\mathfrak{L}\{\bar{f}_{00}, f_{00}\gamma\cdot p + \gamma\cdot p f_{00}\} - m^2\mathfrak{L}\{\bar{f}_{00}, f_{00}\}, \end{aligned}$$

TABLE II. Eigenvalues of nonradial operators in operator set (40) for S–V sector solutions.

Solution type	α	β^2	$\mathbf{J}^2$	J_z	$\bar{\mathcal{I}}$	$\mathcal{R}$
f^{SV}	$4j(j+1)$	0	$J(J+1)$	m	$(-1)^J$	-3
g^{SV+}	$4(j+1)^2$	$4(j+1)^2$	$J(J+1)$	m	$(-1)^{J+1}$	-3
g^{SV-}	$4(j+1)^2$	$4(j+1)^2$	$J(J+1)$	m	$(-1)^J$	-3

$j \in 0, \frac{1}{2}, 1, \frac{3}{2}, \cdots$. Given j, for $f^{SV}J \in 0, 1, 2, \cdots, 2j$ and for $g^{VS\pm}J \in 1, 2, \cdots, 2j+1$. Given J, $m \in J, j-1, \cdots, -J$. The charge parity in each case is $(-1)^{2j}$.

so that

$$(-1)^{2j-J}\mathfrak{L}\{\bar{f}_{00}, -\mathcal{A}_{00}f_{00}\}$$
$$= \int dKK^5\Big(|f_S|^2 - \frac{1}{(2j+1)}(|f_{V1}|^2 - |f_{V2}|^2)$$
$$+ \frac{4[j(j+1)]^{\frac{1}{2}}}{2j+1}\,\mathrm{Re}\,(f_{V1}^* f_{V2})\Big)$$
$$+ 4m\int dKK^4\Big[-\Big(\frac{j+1}{2j+1}\Big)^{\frac{1}{2}}\mathrm{Im}\,(f_S^* f_{V1})$$
$$+ \Big(\frac{j}{2j+1}\Big)^{\frac{1}{2}}\mathrm{Im}\,(f_S^* f_{V2})\Big]$$
$$- m^2\int dKK^3(|f_S|^2 - |f_{V1}|^2 - |f_{V2}|^2), \quad (72)$$

and this can be simplified considerably by using (65) when the potential is due to zero-mass particle exchange.

Next we turn to the evaluation of f_{10} and f_{01} from (50_{10}), and (50_{01}). Since $\lambda_{10} = \lambda_{01} = 0$, these equations become, respectively,

$$(\mathcal{A}_{00} + \lambda_{00}\mathfrak{U})f_{10} = i\{\gamma_4 f_{00}(\gamma\cdot p - im) - (\gamma\cdot p - im)f_{00}\gamma_4\}, \quad (73)$$

$$(\mathcal{A}_{00} + \lambda_{00}\mathfrak{U})f_{01} = -i\{f_{00}\gamma\cdot p - \gamma\cdot p f_{00}\}. \quad (74)$$

If f_{10} or f_{01} is a particular solution of (73) or (74), respectively, a general solution is obtained by adding an arbitrary linear combination of solutions g_i of the homogeneous equation

$$(\mathcal{A}_{00} + \lambda_{00}\mathfrak{U})g_i = 0. \quad (75)$$

Provided λ_{00} is nondegenerate with respect to the quantum numbers $(j^+, j^-, \mathcal{R}')$ or equivalently $(\alpha', \beta'^2, \bar{\mathcal{P}}', \mathcal{R}')$ [see, e.g., Eq. (42)], the only solution of (75) is $g_i = f_{00}$. In this case, a solution (49),

$$f = f_{00} + \epsilon(f_{10} + c_1 f_{00}) + \Delta(f_{01} + c_2 f_{00}) + \underset{r+s=2}{O}(\epsilon^r\Delta^s),$$

can always be recast as

$$f' = f'_{00} + \epsilon f'_{10} + \Delta f'_{01} + \underset{r+s=2}{O}(\epsilon^r\Delta^s),$$

where

$$f' = (1 + \epsilon c_1 + \Delta c_2)^{-1}f,$$
$$f'_{10} = (1 + \epsilon c_1 + \Delta c_2)^{-1}f_{10}, \quad \text{etc.}$$

We can thus arrange without loss of generality that when λ_{00} is nondegenerate, the solution of the homogeneous equation does not contribute to f_{10} or f_{01} in (73) and (74). In the following analysis we shall assume λ_{00} is nondegenerate and that the appropriate arrangement has been made.

If f_{00} was an eigenfunction of $\mathcal{R}$ belonging to eigenvalue $\mathcal{R}'$ say, we saw earlier that $f_{00}\gamma\cdot p - \gamma\cdot p f_{00}$ lay in a sector or sectors orthogonal to f_{00}. The same applies to $\gamma_4 f_{00}(\gamma\cdot p - im) - (\gamma\cdot p - im)f_{00}\gamma_4$. Here, with f_{00} in the S–V sector, the right members of (73) and (74) each lie in the T–A sector. Since from Paper I we know that $\mathcal{A}_{00} + \lambda_{00}\mathfrak{U}$ commutes with $\mathcal{R}$, we thus find from (73) and (74) that both f_{10} and f_{01} belong to the T–A sector.

We now recouple the generalized angular momenta in $|S\rangle$, $|V1\rangle$, and $|V2\rangle$, re-express the results in hyperspherical coordinates (ψ, θ, ϕ), and absorb the ψ-dependent factors into the coefficients. This leads to

$$f_{00} = S_0\,|\Gamma_1 J\,0\,J\,m\rangle + V_1\,|\Gamma_2 J+1\,1\,J\,m\rangle + V_2\,|\Gamma_2 J-1\,1\,J\,m\rangle + V_0\,|\Gamma_2 J\,0\,J\,m\rangle, \quad (76)$$

where

$$|\Gamma_i L\,S\,J\,m\rangle = \sum_{m_l m_s}(L\,S\,m_l\,m_s\,|\,J\,m)Y_{Lm_l}(\theta,\phi)(\Gamma_i)_{Sm_s} \quad (77)$$

and $(\Gamma_i)_{Sm_s}$ is read from Table IV in Appendix B.

The coefficients, with C_n^L defined as in (B4), are

$$S_0 = i^{2j-J}f_S\mathrm{C}_{2j}^J(\psi),$$
$$V_1 = i^{2j-J}\{A_{J+1\,1}f_{V1}\mathrm{C}_{2j+1}^{J+1}(\psi) - B_{J+1\,1}f_{V2}\mathrm{C}_{2j-1}^{J+1}(\psi)\},$$
$$V_2 = i^{2j-J}\{-A_{J-1\,1}f_{V1}\mathrm{C}_{2j+1}^{J-1}(\psi) + B_{J-1\,1}f_{V2}\mathrm{C}_{2j-1}^{J-1}(\psi)\},$$
$$V_0 = i^{2j-J+1}\{A_{J0}f_{V1}\mathrm{C}_{2j+1}^{J}(\psi) - B_{J0}f_{V2}\mathrm{C}_{2j-1}^{J}(\psi)\}, \quad (78)$$

where

$$A_{J\,0} = (2J+1)^{\frac{1}{2}}(2j+1)\begin{Bmatrix} j+\frac{1}{2} & j+\frac{1}{2} & J \\ \frac{1}{2} & \frac{1}{2} & 0 \\ j & j & J \end{Bmatrix},$$

$$A_{J\pm1\,1} = [(2J+3)3]^{\frac{1}{2}}(2j+1) \times \begin{Bmatrix} j+\frac{1}{2} & j+\frac{1}{2} & J\pm1 \\ \frac{1}{2} & \frac{1}{2} & 1 \\ j & j & J \end{Bmatrix}, \quad (79)$$

$$B_{J\,0} = (2J+1)^{\frac{1}{2}}(2j-1)\begin{Bmatrix} j-\frac{1}{2} & j-\frac{1}{2} & J \\ \frac{1}{2} & \frac{1}{2} & 0 \\ j & j & J \end{Bmatrix},$$

$$B_{J\pm1\,1} = [(2J+3)3]^{\frac{1}{2}}(2j-1) \times \begin{Bmatrix} j-\frac{1}{2} & j-\frac{1}{2} & J\pm1 \\ \frac{1}{2} & \frac{1}{2} & 1 \\ j & j & J \end{Bmatrix}.$$

From the properties of $\mathcal{C}_n^L$ under the substitution $\psi \to \pi - \psi$, S_0, V_1, and V_2 have opposite p_4 parity from V_0, e.g., that of S_0 being even or odd according as $2j - J$ is even or odd.

Since the parity of f_{00}, and hence f_{10} and f_{01}, is $(-1)^J$, f_{10} and f_{01} are both of the form

$$g = T_1 |\Gamma_3^- J + 1\ 1\ J\ m\rangle + T_2 |\Gamma_3^- J - 1\ 1\ J\ m\rangle + T_0 |\Gamma_3^+ J\ 1\ J\ m\rangle + A_0 |\Gamma_4 J\ 1\ J\ m\rangle, \quad (80)$$

where T_0, T_1, T_2, and A_0 are each functions of p^2, ψ.

By defining

$$\begin{aligned} a &= \left(\frac{J}{2J+1}\right)^{\frac{1}{2}}, \qquad b = \left(\frac{J+1}{2J+1}\right)^{\frac{1}{2}}, \\ T^+ &= aT_1 + bT_2, \quad V^+ = i(aV_1 + bV_2), \\ T^- &= bT_1 - aT_2, \quad V^- = i(bV_1 - aV_2), \\ U^+ &= T^+ \cos\psi - iT_0 \sin\psi, \\ U^- &= T^+ \sin\psi + iT_0 \cos\psi, \end{aligned} \quad (81)$$

Eq. (73) reduces to

$$(p^2 + m^2)T^- + \lambda_{00}\bar{T}^- = 2pS_0 \sin\psi + 2mV^-, \quad (82a)$$

$$(p^2 + m^2)U^+ + \lambda_{00}\bar{U}^+ = 2mV^+ \cos\psi, \quad (82b)$$

$$2mpA_0 + (p^2 - m^2)U^- - \lambda_{00}\bar{U}^- = -2mV^+ \sin\psi, \quad (82c)$$

$$-2mpU^- + (p^2 - m^2)A_0 - \lambda_{00}\bar{A}_0 = -2pV^+ \sin\psi, \quad (82d)$$

where, for $Q = U^+, U^-, T^-$, or A_0,

$$\bar{Q}(p^2, \psi) = \frac{1}{8\pi^2} \sum_{n=0}^{\infty} \frac{1}{(n+1)} \mathcal{C}_n^L(\psi) \int_0^\infty dK K^3 w_n \times \int_0^\pi d\psi' \sin^2\psi' Q(k^2, \psi') \mathcal{C}_n^L(\psi') \quad (83)$$

and where[14]

$$\begin{aligned} &\mathfrak{W}(|p - k|) \\ &= \sum_{n=0}^{\infty} w_n(k^2, p^2) C_n^1(\cos\chi) \\ &= 2\pi^2 \sum_{n=0}^{\infty} \frac{w_n}{(n+1)} \sum_{l=0}^{n} \sum_{m=-l}^{l} Y^*_{nlm}(\Omega_k) Y_{nlm}(\Omega_p), \end{aligned} \quad (84)$$

for χ the angle between vector p which has direction $\Omega_p = (\psi_p, \theta_p, \phi_p)$ and vector k which has direction $\Omega_k = (\psi_k, \theta_k, \phi_k)$.

Similarly, Eq. (74) reduces to

$$(p^2 + m^2)T^- + \lambda_{00}\bar{T}^- = -2ipV_0 \sin\psi + 2ipV^- \cos\psi, \quad (85a)$$

$$(p^2 + m^2)U^+ + \lambda_{00}\bar{U}^+ = 2ipV^+, \quad (85b)$$

$$2mpA_0 + (p^2 - m^2)U^- - \lambda_{00}\bar{U}^- = 0, \quad (85c)$$

$$-2mpU^- + (p^2 - m^2)A_0 - \lambda_{00}\bar{A}_0 = 0. \quad (85d)$$

We notice that, in each of the equation sets (82) and (85), the equation for T^- is not coupled to the other three.

One could further remove the ψ variables and obtain infinite sets of coupled radial equations.

Throughout this section the $j = 0$ case can be obtained by first setting $f_{V2} = 0$ and afterwards in the remaining elements of the relevant equations setting $j = 0$.

VI. A SPECIAL CASE

A computer program has been developed to test the cases $j = 0$ in order to find eigenvalues of the coupling constant for zero-total-mass systems with amplitudes of type f^{SV}, where the potential is due to exchange of a massless boson. Such systems will have positive parity, and, since $j = 0$ implies $J = 0$, they will also have positive charge parity. The ranges of λ for which eigenvalues are relevant have been determined in Paper III. In Table III below, we list the type of interaction, the theoretically allowed range of λ, the range tested by computer, and the eigenvalue found.

The eigenfunction belonging to the eigenvalue $\lambda_{00} = \frac{1.5}{2}$ is

$$\begin{aligned} 2^{\frac{3}{2}}\pi f_{00}(k) &= f_S(K) - f_{V1}(K)(\gamma \cdot k)/K \\ &= s(K) + v_1(K)(\gamma \cdot k)/K. \end{aligned} \quad (86)$$

s and $(m/K)v_1$ are each polynomials in the reciprocal of the variable

$$y = 1 + K^2/m^2, \quad (87)$$

namely

$$\begin{aligned} s &= 2Ny^{-7}(1 - 14y^{-1} + 56y^{-2} - 84y^{-3} + 42y^{-4}), \\ v_1 &= 7iN(K/m)y^{-8}(1 - 2y^{-1})(1 - 6y^{-1} + 6y^{-2}), \end{aligned} \quad (88)$$

TABLE III. Results of a computer test of allowed ranges of λ.

Type of massless exchange particle	Allowed range of λ	Range tested by computer	Eigenvalue found
S	$0 < \lambda$	$0 < \lambda \leq 10$	nil
V	$0 < \lambda < \frac{1}{16}$	complete allowed range	nil
A	$0 < \lambda$	$0 < \lambda \leq 50$	$\lambda_{00} = \frac{1.5}{2}$
P	$0 < \lambda < \frac{1}{4}$	complete allowed range	nil

where N is a normalization constant that is irrelevant to our discussion.

$$\begin{aligned} &\text{As} \quad K \to 0, \quad s = O(K^0), \quad v_1 = O(K), \\ &\text{as} \quad K \to \infty, \quad s = O(K^{-14}), \quad v_1 = O(K^{-15}). \end{aligned} \tag{89}$$

Conditions (III.I5b) are clearly satisfied.

From (70), (71), and (72), noting that $f_{V2} = 0$, and putting $N = 1$, we obtain

$$\begin{aligned} \mathfrak{L}\{\bar{f}_{00}, f_{00}\} &= \tfrac{1}{2}m^4 \int_1^\infty dy(y-1)(|s|^2 - |v_1|^2) \\ &= -\tfrac{1}{2}m^4 \cdot \frac{2 \cdot 7}{5 \cdot 13 \cdot 17 \cdot 19}, \end{aligned} \tag{90}$$

$$\begin{aligned} \mathfrak{L}\{\bar{f}_{00}, \gamma_4 f_{00} \gamma_4\} &= \tfrac{1}{2}m^4 \int_1^\infty dy(y-1)(|s|^2 + \tfrac{1}{2}|v_1|^2) \\ &= \tfrac{1}{2}m^4 \cdot \frac{2 \cdot 7}{13 \cdot 17 \cdot 19}, \end{aligned} \tag{91}$$

$$\begin{aligned} -\mathfrak{L}\{\bar{f}_{00}, \mathcal{A}_{00} f_{00}\} &= \tfrac{1}{2}m^6 \int_1^\infty dy(y-1)y(|s|^2 - |v_1|^2) \\ &= -\tfrac{1}{2}m^6 \cdot \frac{3 \cdot 7^2}{2^2 \cdot 11 \cdot 13 \cdot 17 \cdot 19}. \end{aligned} \tag{92}$$

Since f_{10} and f_{01} have positive parity, they must each be of the form $T_1 |\Gamma_3^- 1100\rangle$, and, in fact,

$$2^{\frac{3}{2}}\pi f_{10} = -y^{-1}[s - i(m/K)v_1], \tag{93}$$

$$f_{01} = 0. \tag{94}$$

Using (93), we then find

$$\begin{aligned} &i\mathfrak{L}\{\bar{f}_{00}, \gamma_4 f_{10}(\gamma \cdot k - im) - (\gamma \cdot k - im) f_{10} \gamma_4\} \\ &\quad = -\tfrac{1}{2}m^4 \cdot \frac{43 \cdot 947}{2^4 \cdot 3 \cdot 5 \cdot 7 \cdot 13 \cdot 17 \cdot 19}, \end{aligned} \tag{95}$$

and from (62) and (63) we have approximately, for small E and Δ,

$$\lambda \approx \frac{15}{2}\left(1 + 0.766\frac{E^2}{m^2} + 0.838\frac{\Delta^2}{m^2}\right). \tag{96}$$

In summary, when $E/m \ll 1$, $|\Delta|/m \ll 1$, Eq. (96) gives the value for the coupling constant $\lambda = g^2/(4\pi)^2$ that in ladder approximation binds a fermion of mass $m + \Delta$ to an antifermion of mass $m - \Delta$ by exchange of a massless axial vector particle to produce a bound state of mass E having zero angular momentum, positive parity, and positive charge parity.

It is rather disconcerting to find that $\partial\lambda/\partial(E^2)$ is positive, i.e., that the greater the strength of the binding, the greater is the mass of the composite system. Certainly one would insist that to be physically reasonable λ must tend to zero as E approaches $2m$, since this value corresponds to the system becoming unbound. Thus there exists a point between $E = 0$ and $E = 2m$ where either $\partial\lambda/\partial(E^2) = 0$ or else $\partial\lambda/\partial(E^2)$ has a singularity.[15] The result (96) is perhaps also a caution against too readily assuming in other contexts[17] that $\partial\lambda/\partial(E^2)$ is negative.

One feature which has been ignored in our analysis is that, when $\epsilon < \Delta$, the more massive particle (a say) is unstable against the decay $a \to a\bar{b} + b$. This has repercussions, for instance, in the appearance of multiple poles in the scattering Green's function in scalar–scalar equations such as the Wick–Cutkosky model[16] at $\epsilon = \Delta$. We have assumed analyticity of all relevant functions about $\epsilon = \Delta = 0$, in order to establish the framework for the series expansions (47) and (49), and such an assumption may not be justified. However, no pathological features other than the sign of $\partial\lambda/\partial(E^2)$ have been apparent, and we rather doubt that the existence of the decay threshold is the source of this property. Indeed in the ladder approximation the Wick–Cutkosky model exhibits smooth behavior as ϵ passes through the value Δ, as can be seen from Eq. (6.12) of Ref. 16. Questions such as whether the trouble results from the inadequacy of the ladder approximation, of course, lie completely beyond the scope of this investigation.

ACKNOWLEDGMENT

I am indebted to P. W. Johnston for noticing two classes of solution of the ladder approximation Bethe–Salpeter equation for the spin-$\frac{1}{2}$ fermion–antifermion system ($m_a = m_b = m$, $E = 0$) which were overlooked in Paper I (see Appendix E of this paper). These classes have also been obtained by Ito[3] and correspond to his families (iii) and (vii).

APPENDIX A: FURTHER PROPERTIES OF BISPHERICAL ANGULAR FUNCTIONS

The bispherical angular functions $Z_{lm^+m^-}(\nu, \omega, \phi)$ were introduced in Paper I and discussed there and in the Appendix to Paper III. We develop here certain further properties relevant to the present paper.

The rotation operator $D(\alpha\,\beta\,\gamma)$ which rotates a set of coordinate axes through Euler angles (α, β, γ) is defined, e.g., by Edmonds,[18] Eq. (4.1.2). A comparison between the equations which express L^2 and L_z [Edmonds, (2.2.2) and (2.2.3)] in terms of the Euler angles and our equations (I.33) and (I.34) reveals a close similarity. Indeed it is not difficult to show, from Edmonds (4.1.12) and (4.1.23) and our (I.35) and (I.36), that the representative of $D(\omega + \phi, 2\nu, \omega - \phi)$,

$$\begin{aligned} &\mathfrak{D}^{(l)}_{m^+,-m^-}(\omega + \phi, 2\nu, \omega - \phi) \\ &\quad = (-1)^{|m|}\left(\frac{2\pi^2}{2l+1}\right)^{\frac{1}{2}} Z_{lm^+m^-}(\nu, \omega, \phi), \end{aligned} \tag{A1}$$

where $m = m^+ + m^-$.

Applying the formula [cf. Edmonds, Eq. (4.3.1), p. 60]

$$\mathfrak{D}^{(l_1)}_{\alpha\beta}\mathfrak{D}^{(l_2)}_{\gamma\delta} = \sum_j (l_1\, l_2\, \alpha\, \gamma \mid l\, \lambda)(l_1\, l_2\, \beta\, \delta \mid l\, \mu)\mathfrak{D}^{(l)}_{\lambda\mu}, \quad \text{(A2)}$$

we then find as an expansion for the product of two Z functions (of the same arguments ν, ω, ϕ)

$$Z_{lm^+m^-}Z_{kn^+n^-} = \left(\frac{(2l+1)(2k+1)}{2\pi^2}\right)^{\frac{1}{2}} \times \sum_{L=|l-k|}^{l+k} \frac{(-1)^{l+k-L}}{(2L+1)^{\frac{1}{2}}}(l\, k\, m^+\, n^+ \mid L\, M^+) \times (l\, k\, m^-\, n^- \mid L\, M^-)Z_{LM^+M^-}. \quad \text{(A3)}$$

It is readily shown that

$$\int d\Omega Z_{lm^+m^-} \equiv \int_0^{\frac{1}{2}\pi} d\nu \int_0^{2\pi} d\omega \int_0^{2\pi} d\phi \sin\nu\cos\nu Z_{lm^+m^-}(\nu, \omega, \phi) = 2^{\frac{1}{2}}\pi\delta_{l0}\delta_{m^+0}\delta_{m^-0}, \quad \text{(A4)}$$

whence

$$\int d\Omega Z_{lm^+m^-}Z_{kn^+n^-} = (-1)^m\delta_{lk}\delta_{m^+-n^+}\delta_{m^--n^-}. \quad \text{(A5)}$$

Using

$$Z^*_{lm^+m^-} = (-1)^m Z_{l-m^+-m^-}, \quad \text{(A6)}$$

we thus arrive at the orthonormality condition

$$\int d\Omega Z^*_{lm^+m^-}Z_{kn^+n^-} = \delta_{lk}\delta_{m^+n^+}\delta_{m^-n^-}. \quad \text{(A7)}$$

APPENDIX B: RELATION BETWEEN THE BISPHERICAL AND HYPERSPHERICAL BASES IN CONFIGURATION, MOMENTUM, AND DIRAC SPACES

From the relation

$$\mathbf{L} = \mathbf{L}^+ + \mathbf{L}^-, \quad \text{(B1)}$$

the state

$$|(l\, l)L\, M\rangle \equiv \sum_{m^+m^-} (l\, l\, m^+\, m^- \mid L\, M)Z_{lm^+m^-}(\nu, \omega, \phi) \quad \text{(B2)}$$

is a normalized eigenfunction of $(\mathbf{L}^+)^2 \equiv (\mathbf{L}^-)^2$, $\mathbf{L}^2$, L_z.

But the normalized hyperspherical harmonic

$$Y_{nLM}(\psi, \theta, \phi) = \mathcal{C}^L_n(\psi)Y_{LM}(\theta, \phi), \quad \text{(B3)}$$

$$\mathcal{C}^L_n(\psi) = \left(\frac{2^{2L+1}}{\pi}\cdot\frac{(n+1)(n-L)!}{(n+L+1)!}\right)^{\frac{1}{2}} \times L!\sin^L\psi C^{L+1}_{n-L}(\cos\psi), \quad \text{(B4)}$$

where Y_{LM} is the usual spherical harmonic and C^{L+1}_{n-L} is a Gegenbauer polynomial, is an eigenfunction of $\alpha_L = 2\{(\mathbf{L}^+)^2 + (\mathbf{L}^-)^2\}$, $\mathbf{L}^2$ and L_z belonging, respectively, to eigenvalues $n(n+2)$, $L(L+1)$, and M [cf. (I.26) *et seq.*]. Thus, $|(l\, l)L\, M\rangle$ may also be represented by Y_{2lLM}.

The phases of the two representations may differ and so we put

$$\sum_{m^+m^-} (l\, l\, m^+\, m^- \mid L\, M)Z_{lm^+m^-} = e^{i\zeta}Y_{2lLM}, \quad \text{(B5)}$$

where $\zeta = \zeta(l\, L\, M)$ and is real. From $\mathbf{L} = \mathbf{L}^+ + \mathbf{L}^-$ we have for the step operators $L_\pm = L_x \pm iL_y$, etc.,

$$L_\pm = L^+_\pm + L^-_\pm, \quad \text{(B6)}$$

and, applying this to (B5), we find that ζ is independent of M for given l and L. From the definitions of x_μ in terms of bispherical coordinates and in terms of hyperspherical coordinates, putting $\nu = 0$, we find is equivalent to putting $\theta = 0$ and $\omega = \frac{1}{2}\pi - \psi$ (mod 2π) or putting $\theta = \pi$ and $\omega = \frac{1}{2}\pi + \psi$. If we substitute these values into (B5), it is not difficult to show that $\zeta = \frac{1}{2}\pi(2l - L)$, so that

$$\sum_{m^+m^-} (l\, l\, m^+\, m^- \mid L\, M)Z_{lm^+m^-} = e^{i\frac{1}{2}\pi(2l-L)}Y_{2lLM}. \quad \text{(B7)}$$

Exactly the same relation obtains in the 4-dimensional Euclidean relative momentum space, since in place of (I.7) we have

$$\mathcal{M}_{\mu\nu} = -i(p_\mu\partial_\nu - p_\nu\partial_\mu) + \tfrac{1}{2}[\sigma_{\mu\nu}, \ ], \quad \text{(B8)}$$

which is equivalent formally to replacing $x_\mu \to p_\mu$ in all generalized angular momentum operators, and the functions $Z_{lm^+m^-}$ and Y_{2lLM} have exactly the same interpretation as they do in relative configuration space.

If we take the Dirac-space eigenfunctions

$$|s^+\, s^-\, m^+_s\, m^-_s\rangle$$

to be as given in Table II of Paper I, then the corresponding spherical basis eigenfunctions defined in the standard way are

$$|s^+\, s^-\, S\, m_s\rangle = \sum_{m_s^+m_s^-} (s^+\, s^-\, m^+_s\, m^-_s \mid S\, m_s)\,|s^+\, s^-\, m^+_s\, m^-_s\rangle. \quad \text{(B9)}$$

It is convenient, however, in connection with the parity operation, to ensure that the matrices chosen are eigenfunctions of $\hat{\gamma}_4$. This is already the case for those of type Γ_1, Γ_2, Γ_4, and Γ_5, but not for the two sets of type Γ_3. We overcome the problem by taking linear combinations of the two sets to give two new sets Γ^+_3 and Γ^-_3 which satisfy $\hat{\gamma}_4\Gamma^\pm_3 = \pm\Gamma^\pm_3$. With this modification the spherical basis eigenfunctions are listed in Table IV below. In each case a normalization factor $\frac{1}{2}$ is required.

TABLE IV. Spherical basis Dirac-space eigenfunctions.

	S	m_s	Eigenfunctions (not normalized)	
$s^+ = s^- = 0$	0	0	I	γ_5
$s^+ = s^- = \frac{1}{2}$	0	0	$i\gamma_4$	$i \cdot i\gamma_5\gamma_4$
	1	1	$-2^{-\frac{1}{2}}(\gamma_1 + i\gamma_2)$	$-2^{-\frac{1}{2}}i\gamma_5(\gamma_1 + \gamma_2)$
	1	0	γ_3	$i\gamma_5\gamma_3$
	1	−1	$2^{-\frac{1}{2}}(\gamma_1 - i\gamma_2)$	$2^{-\frac{1}{2}}i\gamma_5(\gamma_1 - i\gamma_2)$
Linear combination				
$s^+ = 1, s^- = 0$	1	1	Γ_3^-	Γ_3^+
$s^+ = 0, s^- = 1$	1	0	$-2^{-\frac{1}{2}}(\sigma_{41} + i\sigma_{42})$	$-2^{-\frac{1}{2}}(\sigma_{23} + i\sigma_{31})$
	1	−1	σ_{43}	σ_{12}
			$2^{-\frac{1}{2}}(\sigma_{41} - i\sigma_{42})$	$2^{-\frac{1}{2}}(\sigma_{23} - i\sigma_{31})$

APPENDIX C: TENSOR OPERATORS FOR ROTATIONS IN A FOUR-DIMENSIONAL EUCLIDEAN SPACE

Let α be a 4×4 matrix whose elements may be functions of spatial variables. As a premultiplier, it is a linear operator when applied to arbitrary 4×4 matrices. Denoting it by $(a \times I)$ and using the bra and ket notation in the same sense as in Sec. III, we may write

$$(\alpha \times I)\,|A\rangle = \alpha A(v, \omega, \phi). \tag{C1}$$

Similarly with α as a postmultiplier, we have the linear operator $(I \times \alpha)$:

$$(I \times \alpha)\,|A\rangle = A(v, \omega, \phi)\alpha. \tag{C2}$$

More generally we have linear operators of the form $(\alpha \times \beta)$, with

$$(\alpha \times \beta)\,|A\rangle = \alpha A \beta, \tag{C3}$$

and such relations as

$$(\alpha \times \beta)(\gamma \times \delta) = (\alpha\gamma \times \delta\beta). \tag{C4}$$

By requiring, in the usual way, that an operator be associative in an arbitrary scalar product, we have

$$\begin{aligned}\{\langle B|\,(\alpha \times \beta)\}\,|A\rangle &= \langle B|\,\{(\alpha \times \beta)\,|A\rangle\} \\ &= \int d\Omega\, \mathrm{Tr}\,(B^\dagger \alpha A \beta) \\ &= \int d\Omega\, \mathrm{Tr}\,(\beta B^\dagger \alpha A)\end{aligned} \tag{C5}$$

for all $\langle B|$ and $|A\rangle$, so that

$$\langle B|\,(\alpha \times \beta) = \beta B^\dagger \alpha. \tag{C6}$$

We also find

$$(\alpha \times \beta)^\dagger = (\alpha^\dagger \times \beta^\dagger). \tag{C7}$$

Finally, we notice the duality which exists between linear operators and vectors:

If the matrices α and β are allowed forms for the 4×4 matrices that represent vectors $|\alpha\rangle$ and $|\beta\rangle$, respectively, then

$$\begin{aligned}\langle B|\,(\alpha \times \beta^\dagger)\,|A\rangle &= \int d\Omega\, \mathrm{Tr}\,(B^\dagger \alpha A \beta^\dagger) \\ &= \int d\Omega\, \mathrm{Tr}\,(\beta^\dagger B^\dagger \alpha A) \\ &= \langle \beta|\,(B^\dagger \times A)\,|\alpha\rangle.\end{aligned} \tag{C8}$$

Under an arbitrary infinitesimal rotation of the physical system determined by real parameters $\alpha_{\mu\nu}$ (with $\alpha_{\nu\mu} = -\alpha_{\mu\nu}$), the 2-body amplitude χ becomes

$$\chi' = (1 - i\alpha_{\mu\nu}\mathcal{M}_{\mu\nu})\chi \tag{C9}$$

and similarly

$$\chi^{\dagger\prime} = \chi'^\dagger = (1 - i\alpha_{\mu\nu}\mathcal{M}_{\mu\nu})\chi^\dagger. \tag{C10}$$

In the center-of-mass system and when $E = 0$, we have [cf. (I.7)] in configuration space

$$\mathcal{M}_{\mu\nu} = -i(x_\mu\partial_\nu - x_\nu\partial_\mu) + \tfrac{1}{2}[\sigma_{\mu\nu}, \quad] \tag{C11}$$

and in relative momentum space

$$\mathcal{M}_{\mu\nu} = -i(p_\mu\partial'_\nu - p_\nu\partial'_\mu) + \tfrac{1}{2}[\sigma_{\mu\nu}, \quad], \tag{C12}$$

where $\partial_\mu = \partial/\partial x_\mu$, $\partial'_\mu = \partial/\partial p_\mu$.

Radial factors are clearly irrelevant so far as such rotations are concerned and so we shall concentrate on just the angular and Dirac-space factors.

Let us require that under arbitrary rotations scalar products

$$\langle A|\,(\alpha \times \beta)\,|B\rangle = \int d\Omega\, \mathrm{Tr}\,(A^\dagger \alpha B \beta), \tag{C13}$$

such as we have been considering, remain invariant. From this we now deduce that each element ϕ in the scalar product (i.e., $\phi = A$, α, B, β) undergoes the same transformation:

$$\phi \to \phi' = (1 - i\alpha_{\mu\nu}\mathcal{M}_{\mu\nu})\phi. \tag{C14}$$

Under the infinitesimal unitary transformation

$$U = \exp(-i\alpha_{\mu\nu}\mathcal{M}_{\mu\nu}) = 1 - i\alpha_{\mu\nu}\mathcal{M}_{\mu\nu}, \tag{C15}$$

kets, bras, and linear operators transform, respectively, as

$$|A\rangle \to |A\rangle' = U|A\rangle = (1 - i\alpha_{\mu\nu}\mathcal{M}_{\mu\nu})|A\rangle, \quad \text{(C16)}$$

$$\langle B| \to \langle B|' = \langle B|U^\dagger = \langle B|(1 + i\alpha_{\mu\nu}\mathcal{M}_{\mu\nu}), \quad \text{(C17)}$$

$$\beta \to \beta' = U\beta U^\dagger = \beta - i\alpha_{\mu\nu}[\mathcal{M}_{\mu\nu}, \beta]_{\text{op}}, \quad \text{(C18)}$$

where we demonstrate the significance of the suffix "op" as follows:

$$\begin{aligned}\beta'|A\rangle' = U\beta|A\rangle &= (1 - i\alpha_{\mu\nu}\mathcal{M}_{\mu\nu})\beta|A\rangle \\ &= \beta|A\rangle - i\alpha_{\mu\nu}\mathcal{M}_{\mu\nu}(\beta|A\rangle) \\ &= \beta(|A\rangle' + i\alpha_{\mu\nu}\mathcal{M}_{\mu\nu}|A\rangle') \\ &\quad - i\alpha_{\mu\nu}\mathcal{M}_{\mu\nu}(\beta|A\rangle') \\ &= (\beta - i\alpha_{\mu\nu}[\mathcal{M}_{\mu\nu}, \beta]_{\text{op}})|A\rangle'. \quad \text{(C19)}\end{aligned}$$

In this it is clear that the $\mathcal{M}_{\mu\nu}$ within the commutator acts not only on β but also on the ket $|A\rangle'$ to which β is applied.

But also

$$\begin{aligned}\beta'|A\rangle' &= (1 - i\alpha_{\mu\nu}\mathcal{M}_{\mu\nu})\beta|A\rangle \\ &= \{1 - i\alpha_{\mu\nu}(\mathcal{M}_{\mu\nu}\beta)|A\rangle - i\beta\alpha_{\mu\nu}(\mathcal{M}_{\mu\nu}|A\rangle)\} \\ &= \{(1 - i\alpha_{\mu\nu}\mathcal{M}_{\mu\nu})\beta\}\{(1 - i\alpha_{\mu\nu}\mathcal{M}_{\mu\nu})|A\rangle\} \\ &= \{(1 - i\alpha_{\mu\nu}\mathcal{M}_{\mu\nu})\beta\}|A\rangle', \quad \text{(C20)}\end{aligned}$$

so that

$$\beta' = (1 - i\alpha_{\mu\nu}\mathcal{M}_{\mu\nu})\beta, \quad \text{(C21)}$$

where we have omitted the suffix "op," since here the $\mathcal{M}_{\mu\nu}$ acts only on β. For an outer product operator $(\alpha \times \beta)$ we clearly have

$$(\alpha \times \beta)' = (U\alpha U^\dagger \times U\beta U^\dagger), \quad \text{(C22)}$$

so that with this and (C21) we have verified (C14). This result further emphasizes the duality that exists between the vectors and linear operators.

Let us now change the basis of our rotation generators and write

$$\alpha_{\mu\nu}\mathcal{M}_{\mu\nu} = \boldsymbol{\alpha}^+ \cdot \mathbf{J}^+ + \boldsymbol{\alpha}^- \cdot \mathbf{J}^-, \quad \text{(C23)}$$

where $\mathbf{J}^\pm$ are defined in (I.29). Following standard procedure (e.g., Brink and Satchler,[10] Chap. 4.5, p. 52 *et seq.*) we define the components of an irreducible tensor operator $\beta^\pm_{kq}$ in the $N^\pm_{3P}$ subspace by

$$\begin{aligned}[J_z^\pm, \beta^\pm_{kq}]_{\text{op}} &= q\beta^\pm_{kq}, \\ [J_+^\pm, \beta^\pm_{kq}]_{\text{op}} &= [(k + q + 1)(k - q)]^{\frac{1}{2}}\beta^\pm_{kq+1}, \quad \text{(C24)} \\ [J_-^\pm, \beta^\pm_{kq}]_{\text{op}} &= [(k - q + 1)(k + q)]^{\frac{1}{2}}\beta^\pm_{kq-1},\end{aligned}$$

from which also follows

$$\sum_i [J_i^\pm, [J_i^\pm, \beta^\pm_{kq}]_{\text{op}}]_{\text{op}} = k(k + 1)\beta^\pm_{kq}. \quad \text{(C25)}$$

Comparing (C19) and (C20), we see that these may also be written

$$\begin{aligned}J_z^\pm\beta^\pm_{kq} &= q\beta^\pm_{kq}, \\ J_+^\pm\beta^\pm_{kq} &= [(k + q + 1)(k - q)]^{\frac{1}{2}}\beta^\pm_{kq+1}, \quad \text{(C26)} \\ J_-^\pm\beta^\pm_{kq} &= [(k - q + 1)(k + q)]^{\frac{1}{2}}\beta^\pm_{kq-1},\end{aligned}$$

$$(\mathbf{J}^\pm)^2\beta^\pm_{kq} = k(k + 1)\beta^\pm_{kq}. \quad \text{(C27)}$$

These last results imply that irreducible tensor operator components of the form $\beta^\pm_{kq}(p)$ have analogous matrix expression to the eigenfunctions belonging to eigenvalues q and $k(k + 1)$ for $J_z^\pm$ and $(\mathbf{J}^\pm)^2$, respectively. In particular, if $\beta_{k+k-q+q-}(p)$ is a component of an irreducible tensor operator of rank k^+ in subspace N^+_{3P} and rank k^- in subspace N^-_{3P}, then

$$\beta_{k+k-q+q-}(p) = \beta(|p|) \times \text{matrix of } |k^+k^-q^+q^-\rangle, \quad \text{(C28)}$$

where $\beta(|p|)$ is a scalar function of $|p|$.

$$\beta_{kq} = \sum_{q^+q^-} (k^+k^-q^+q^- \mid kq)\beta_{k+k-q+q-} \quad \text{(C29)}$$

is the q component of an irreducible tensor operator of rank k with respect to ordinary 3-dimensional rotations.

A linear operator of the form $(\beta_{k+k-q+q-} \times \gamma_{l+l-r+r-})$ transforms as a tensor component of the product of two irreducible representations.

Similarly,

$$\begin{aligned}&\sum_{q^+q^-r^+r^-Q^+Q^-} (K^+K^-Q^+Q^- \mid KQ)(k^+l^+q^+r^+ \mid K^+Q^+) \\ &\quad \times (k^-l^-q^-r^- \mid K^-Q^-)(\beta_{k+k-q+q-} \times \gamma_{l+l-r+r-}) \quad \text{(C30)}\end{aligned}$$

transforms as the Q component of a tensor operator of rank K under ordinary 3-dimensional rotations.

APPENDIX D: REDUCTION OF MATRIX ELEMENTS

The general type of matrix element to be reduced is

$$\langle \Gamma_i(l^+ s^+)j^+(l^- s^-)j^-; J\, m|\, \mathcal{O}\, |\Gamma_j(l_1^+ s_1^+)j_1^+(l_1^- s_1^-)j_1^-; J\, m\rangle, \quad \text{(D1)}$$

where $\mathcal{O}$ is a scalar under 4D rotations. If one could split the terms of $\mathcal{O}$ into factors referring separately to the N^+_{3P}, N^-_{3P} subspaces, the reduction of the matrix element would be simple. Unfortunately this is not possible, and we can only express the terms of $\mathcal{O}$ as scalar products of tensor operators in momentum (or configuration) space and in Dirac space. We then have to recouple the generalized angular momenta into ordinary orbital and spin-parts in the bra and the ket. Thus,

$$\mathcal{O} = \sum_a R^a_k \cdot S^a_k, \quad \text{(D2)}$$

where R_k^a is an irreducible tensor operator of rank k in momentum space and S_k^a is an irreducible tensor operator of rank k in Dirac space. We have

$$\langle \Gamma_i(l^+ s^+)j^+(l^- s^-)j^-; J\, m| = \sum_{LS} [(2L+1)(2S+1)(2j^+ + 1)(2j^- + 1)]^{\frac{1}{2}} \times \begin{Bmatrix} l^+ & l^- & L \\ s^+ & s^- & S \\ j^+ & j^- & J \end{Bmatrix} \langle \Gamma_i(l^+ l^-)L(s^+ s^-)S; J\, m|, \quad \text{(D3)}$$

and there is a similar expansion for the ket. The matrix element (D1) can now be expressed as a linear combination of terms of type

$$\langle \Gamma_i(l^+ l^-)L(s^+ s^-)S; J\, m| R_k^a \cdot S_k^a |\Gamma_j(l_1^+ l_1^-)L'(s_1^+ s_1^-)S'; J\, m\rangle = [(2L+1)(2S+1)]^{\frac{1}{2}}(-1)^{J-L-S'} W(LL'SS'; kJ) \times \langle (l^+ l^-)L| |R_k^a| |(l_1^+ l_1^-)L'\rangle \times \langle \Gamma_i(s^+ s^-)S| |S_k^a| |\Gamma_j(s_1^+ s_1^-)S'\rangle, \quad \text{(D4)}$$

where the reduced matrix elements are defined as in Eq. (4.15) of Ref. 10. The final step is then to determine the relevant reduced matrix elements for substitution into (D4).

Consider first the reduction of

$$\mathfrak{L}\{\bar{f}_{00}, \gamma_4 f_{00}\gamma_4\} = \int dK K^3 F(K), \quad \text{(D5)}$$

where

$$F(K) = (\langle S| f_S^* - \langle V1| f_{V1}^* - \langle V2| f_{V2}^*) \times (\gamma_4 \times \gamma_4)(f_S |S\rangle + f_{V1} |V1\rangle + f_{V2} |V2\rangle) \quad \text{(D6)}$$

and where $K = |p|$. We can reduce the labor by noticing that because of the trace operation in the scalar product all cross terms between S and V terms vanish, and because of the angular integration the V_1–V_2 cross terms vanish. Therefore,

$$F(K) = |f_S|^2 \langle S| (\gamma_4 \times \gamma_4) |S\rangle - |f_{V1}|^2 \langle V1| (\gamma_4 \times \gamma_4) |V1\rangle - |f_{V2}|^2 \langle V2| (\gamma_4 \times \gamma_4) |V2\rangle. \quad \text{(D7)}$$

From Table IV, (C28) *et seq.*, we see that $(\gamma_4 \times \gamma_4)$ is an irreducible tensor operator of rank 0 in Dirac space with respect to 3D rotations. We then require the reduced matrix elements

$$\begin{aligned} &\langle \Gamma_1(0\,0)\,0| |(\gamma_4 \times \gamma_4)| |\Gamma_1(0\,0)\,0\rangle = 1, \\ &\langle \Gamma_2(\tfrac{1}{2}\,\tfrac{1}{2})\,0| |(\gamma_4 \times \gamma_4)| |\Gamma_2(\tfrac{1}{2}\,\tfrac{1}{2})\,0\rangle = 1, \\ &\langle \Gamma_2(\tfrac{1}{2}\,\tfrac{1}{2})\,1| |(\gamma_4 \times \gamma_4)| |\Gamma_2(\tfrac{1}{2}\,\tfrac{1}{2})\,1\rangle = -1. \end{aligned} \quad \text{(D8)}$$

Using these with (D3) (and its ket analog), (D4), and (D7), we obtain

$$F(K) = |f_S|^2 + |f_{V1}|^2 \left(\frac{2(j+1) + J(J+1)}{4(j+1)^2}\right) - |f_{V2}|^2 \left(\frac{2j - J(J+1)}{4j^2}\right). \quad \text{(D9)}$$

Next consider

$$\mathfrak{L}\{\bar{f}_{00}, -\mathcal{A}_{00} f_{00}\} = \int d^4p \,\mathrm{Tr}\,\{\bar{f}_{00}[\gamma \cdot p f_{00}\gamma \cdot p - im(f_{00}\gamma \cdot p + \gamma \cdot p f_{00}) - m^2 f_{00}]\}. \quad \text{(D10)}$$

The last term is determined at once from (70). From (55),

$$\mathrm{Tr}\,(\bar{f}_{00} f_{00}\gamma \cdot p) = \mathrm{Tr}\,(\bar{f}_{00}\gamma \cdot p f_{00}). \quad \text{(D11)}$$

Thus

$$\int d^4p \,\mathrm{Tr}\,[\bar{f}_{00}(f_{00}\gamma \cdot p + \gamma \cdot p f_{00})] = 2\int dK K^3 G(K), \quad \text{(D12)}$$

where

$$G(K) = \int d\Omega \,\mathrm{Tr}\,(\bar{f}_{00} f_{00}\gamma \cdot p) = (\langle S| f_S^* - \langle V1| f_{V1}^* - \langle V2| f_{V2}^*)(I \times \gamma \cdot p) \times (f_S |S\rangle + f_{V1} |V1\rangle + f_{V2} |V2\rangle). \quad \text{(D13)}$$

This time we see that *only* cross products between S and V vectors contribute, and

$$G(K) = 2i\,\mathrm{Im}\,\{f_S^* f_{V1} \langle S| (I \times \gamma \cdot p) |V1\rangle + f_S^* f_{V2} \langle S| (I \times \gamma \cdot p) |V2\rangle\}. \quad \text{(D14)}$$

For $(I \times \gamma \cdot p)$ the expansion (D2) becomes

$$(I \times \gamma \cdot p) = p_4(I \times \gamma_4) + \mathbf{p} \cdot (I \times \boldsymbol{\gamma}), \quad \text{(D15)}$$

and the relevant reduced matrix elements are

$$\begin{aligned} &\langle (j\,j)J| |p_4| |(j+\tfrac{1}{2}\,j+\tfrac{1}{2})J\rangle = \tfrac{1}{2}iK\left(\frac{(2j+J+2)(2j-J+1)}{(2j+1)(2j+2)}\right)^{\frac{1}{2}}, \\ &\langle (j\,j)J| |p_4| |(j-\tfrac{1}{2}\,j-\tfrac{1}{2})J\rangle = -\tfrac{1}{2}iK\left(\frac{(2j-J)(2j+J+1)}{2j(2j+1)}\right)^{\frac{1}{2}}, \\ &\langle (j\,j)J| |\mathbf{p}| |(j+\tfrac{1}{2}\,j+\tfrac{1}{2})J+1\rangle = -\tfrac{1}{2}K\left(\frac{J+1}{2J+1}\cdot\frac{(2j+J+3)(2j+J+2)}{(2j+1)(2j+2)}\right)^{\frac{1}{2}}, \\ &\langle (j\,j)J| |\mathbf{p}| |(j-\tfrac{1}{2}\,j-\tfrac{1}{2})J+1\rangle = -\tfrac{1}{2}K\left(\frac{J+1}{2J+1}\cdot\frac{(2j-J)(2j-J-1)}{2j(2j+1)}\right)^{\frac{1}{2}}, \\ &\langle (j\,j)J| |\mathbf{p}| |(j+\tfrac{1}{2}\,j+\tfrac{1}{2})J-1\rangle = \tfrac{1}{2}K\left(\frac{J}{2J+1}\cdot\frac{(2j-J+2)(2j-J+1)}{(2j+1)(2j+2)}\right)^{\frac{1}{2}}, \\ &\langle (j\,j)J| |\mathbf{p}| |(j-\tfrac{1}{2}\,j-\tfrac{1}{2})J-1\rangle = \tfrac{1}{2}K\left(\frac{J}{2J+1}\cdot\frac{(2j+J+1)(2j+J)}{2j(2j+1)}\right)^{\frac{1}{2}}, \end{aligned} \quad \text{(D16)}$$

$$\langle \Gamma_1(0\,0)0| |(I \times \gamma_4)| |\Gamma_2(\tfrac{1}{2}\,\tfrac{1}{2})0\rangle = i,$$

$$\langle \Gamma_1(0\,0)0| |(I \times \boldsymbol{\gamma})| |\Gamma_2(\tfrac{1}{2}\,\tfrac{1}{2})1\rangle = -3^{\frac{1}{2}},$$

whence

$$\langle S|(I\times\gamma\cdot p)|V1\rangle = -K\left(\frac{j+1}{2j+1}\right)^{\frac{1}{2}},$$
$$\langle S|(I\times\gamma\cdot p)|V2\rangle = K\left(\frac{j}{2j+1}\right)^{\frac{1}{2}}, \quad \text{(D17)}$$

and

$$\int d^4p\,\mathrm{Tr}\,[\bar{f}_{00}(f_{00}\gamma\cdot p+\gamma\cdot p f_{00})]$$
$$= 4i\int dKK^4\left[-\left(\frac{j+1}{2j+1}\right)^{\frac{1}{2}}\mathrm{Im}\,(f_S^* f_{V1})\right.$$
$$\left.+\left(\frac{j}{2j+1}\right)^{\frac{1}{2}}\mathrm{Im}\,(f_S^* f_{V2})\right]. \quad \text{(D18)}$$

Finally

$$\int d^4p\,\mathrm{Tr}\,(\bar{f}_{00}\gamma\cdot p f_{00}\gamma\cdot p)=\int dKK^3H(K),$$
$$H(K)=|f_S|^2\langle S|(\gamma\cdot p\times\gamma\cdot p)|V2\rangle$$
$$-|f_{V1}|^2\langle V1|(\gamma\cdot p\times\gamma\cdot p)|V1\rangle$$
$$-|f_{V2}|^2\langle V2|(\gamma\cdot p\times\gamma\cdot p)|V2\rangle$$
$$-2\,\mathrm{Re}\,[f_{V1}^* f_{V2}\langle V1|(\gamma\cdot p\times\gamma\cdot p)|V2\rangle]. \quad \text{(D19)}$$

The expansion (D2) in this case becomes

$$(\gamma\cdot p\times\gamma\cdot p)=p_4^2(\gamma_4\times\gamma_4)+p_4\mathbf{p}\cdot(\gamma_4\times\boldsymbol{\gamma})$$
$$+p_4\mathbf{p}\cdot(\boldsymbol{\gamma}\times\gamma_4)+\tfrac{1}{3}\mathbf{p}^2(\gamma_i\times\gamma_i)$$
$$+(p_ip_j-\tfrac{1}{3}\mathbf{p}^2\delta_{ij})\cdot(\gamma_i\times\gamma_j). \quad \text{(D20)}$$

A tedious calculation then gives

$$\int d^4p\,\mathrm{Tr}\,(\bar{f}_{00}\gamma\cdot p f_{00}\gamma\cdot p)$$
$$=\int dKK^5\left(|f_S|^2-\frac{1}{(2j+1)}(|f_{V1}|^2-|f_{V2}|^2)\right.$$
$$\left.+\frac{4[j(j+1)]^{\frac{1}{2}}}{2j+1}\,\mathrm{Re}\,(f_{V1}^* f_{V2})\right). \quad \text{(D21)}$$

APPENDIX E: FURTHER S–V SECTOR SOLUTIONS

If one assumes the existence of S–V sector solutions with the S term vanishing identically, we have (following Sec. V of Paper I) the possibilities

$$\text{(B')}\quad j^+=j^--1=j,\quad l_V=j+\tfrac{1}{2},$$
$$\text{(C')}\quad j^+-1=j^-=j,\quad l_V=j+\tfrac{1}{2} \quad \text{(E1)}$$

leading to respective solutions

$$f^{SV}_{j\,j+1\,m^+m^-}=f_V|(j+\tfrac{1}{2}\,\tfrac{1}{2})j(j+\tfrac{1}{2}\,\tfrac{1}{2})j+1;m^+m^-\rangle,$$
$$f^{SV}_{j+1\,j\,m^+m^-}=f_V|(j+\tfrac{1}{2}\,\tfrac{1}{2})j+1(j+\tfrac{1}{2}\,\tfrac{1}{2})j;m^+m^-\rangle, \quad \text{(E2)}$$

where in each case the radial function $v=f_V(R)$ satisfies

$$\begin{Bmatrix}D_2^+D_{\frac{1}{2}}^-\\ D_0^-D_{\frac{3}{2}}^+\end{Bmatrix}v=(m^2+\lambda\bar{V}_2)v. \quad \text{(E3)}$$

From the Appendix of Paper I we see that both solutions are "transverse" since

$$\Delta'\cdot V=0. \quad \text{(E4)}$$

Angular momentum eigenstates are obtained as in Eq. (II.9) by coupling:

$$f^{SV}_{j\,j+1\,J\,m}=f_V|(j+\tfrac{1}{2}\,\tfrac{1}{2})j(j+\tfrac{1}{2}\,\tfrac{1}{2})j+1;J\,m\rangle,\quad J\geq 1,$$
$$f^{SV}_{j+1\,j\,J\,m}=f_V|(j+\tfrac{1}{2}\,\tfrac{1}{2})j+1(j+\tfrac{1}{2}\,\tfrac{1}{2})j;J\,m\rangle,\quad J\geq 1 \quad \text{(E5)}$$

and parity eigenstates by taking the linear combinations [cf. (II.12)]

$$g^{SV\pm}_{j\,j+1\,J\,m}=f^{SV}_{j\,j+1\,J\,m}\pm f^{SV}_{j+1\,j\,J\,m}. \quad \text{(E6)}$$

Being in the S–V sector all these functions have charge parity $(-1)^{2j}$.

[1] H. A. Bethe and E. E. Salpeter, Phys. Rev. **84**, 1232 (1951).
[2] M. Gell-Mann and F. E. Low, Phys. Rev. **84**, 350 (1951).
[3] H. Ito, Progr. Theoret. Phys. (Kyoto) **41**, 1109 (1969).
[4] See our Sec. IV.
[5] R. F. Keam, J. Math. Phys. **9**, 1462 (1968). Hereafter referred to as I. Equations from paper I will be notated here as, e.g., Eq. (I.20).
[6] R. F. Keam, J. Math. Phys. **10**, 594 (1969). Hereafter referred to as II. Notation of equations will be analogous to that in Ref. 5.
[7] R. F. Keam, J. Math. Phys. **11**, 394 (1970). Hereafter referred to as III. Notation of equations will be analogous to that in Ref. 5.
[8] S. Naito and N. Nakanishi, Progr. Theoret. Phys. (Kyoto) **42**, 402 (1969).
[9] E.g., P. Roman, *Theory of Elementary Particles* (North-Holland, Amsterdam, 1961), Chap. IV, Sec. 3.
[10] D. M. Brink and G. R. Satchler, *Angular Momentum* (Clarendon, Oxford, 1968).
[11] Note that because of the minus sign in the right member of (26), $\bar{\mathfrak{G}}(Z\Lambda)=-\bar{\mathfrak{G}}(Z)\bar{\mathfrak{G}}(\Lambda)$ in any product of an angular function Z and Dirac-space matrix Λ.
[12] See the discussion of Eq. (42).
[13] See Appendix E.
[14] M. Gourdin, Nuovo Cimento **7**, 338 (1958), Appendix I.
[15] *Note added in proof:* I am grateful to R. H. Dalitz and N. Nakanishi for pointing out a very unpleasant consequence if $\partial\lambda/\partial(E^2)$ vanishes at some point. From Eq. (4.5) of Ref. 16 we see that the normalization constant of the Bethe–Salpeter amplitude would become infinite at this point, so that the pole term of the scattering Green's function can no longer be written as (3.2) of Ref. 16, and this is the fundamental postulate in deriving the homogeneous Bethe–Salpeter equation.
[16] N. Nakanishi, Progr. Theoret. Phys. (Kyoto) Suppl. **43**, 52 (1969).
[17] See, e.g., K. Seto, Progr. Theoret. Phys. (Kyoto) **42**, 1394 (1969), Eq. (4.12) *et seq.*
[18] A. R. Edmonds, *Angular Momentum in Quantum Mechanics* (Princeton U.P., Princeton, N.J., 1957).

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 12, NUMBER 3 MARCH 1971

Coupling of Waves in Nonuniform Plasma

B. CHAKRABORTY
Department of Mathematics, Jadavpur University, Calcutta-32, India

(Received 11 December 1969; Revised Manuscript Received 10 June 1970)

This paper studies mutual coupling of transverse and longitudinal high-frequency waves of both electron and ion fluid motions of perturbation in slowly varying nonmagnetized equilibrium plasma. Besides, the particular problem of coupling of ion acoustic and electromagnetic waves is considered in detail. The power of electromagnetic radiation generated by an ion acoustic wave and also in reverse from the latter to the former have been evaluated.

1. INTRODUCTION

Wave propagation through ionized and other media requiring the use of Maxwell's equations of electrodynamics have gained recently much importance.[1–14] In addition to this very long list, one may refer to papers[15–20] which have discussed wave propagation from a somewhat similar point of view.

Perturbation in uniform infinite plasma travels as transverse and longitudinal waves.[7] If both electron and ion oscillations are included, then two modes of mutually connected longitudinal oscillation are obtained in addition to a mode of transverse field, which propagates independent of the former. Hence no transfer of power occurs from one form to the other in this idealization.

The opposite extreme of this simplification is the limit of most randomly fluctuating medium, where development and transmission of fluctuation has to be considered from an entirely different point of view.[21]

Between these extreme limiting cases lies the intermediate class of problems of current interest, namely wave propagation in a slowly varying medium, where approximations of either case do not apply. A mathematical basis for the study of this class in plasmas has not yet been developed satisfactorily. The plasma equations, even in the approximation of two-component fully ionized fluid mixture (not to speak of the statistical distribution functions in the 6-dimensional phase space), are complicated. In a slowly varying medium the wavelength of propagation has practically an upper limit, which should be much less than the characteristic length of variation of any equilibrium parameter. Then the pressure and the transverse perturbation fields are mutually connected, and so excitation of one leads to the growth of the other. This results in mutual power transfer.[15,16] The wave equation for the magnetic field vector contains source terms depending on the pressure perturbation, and similarly the wave equation of the latter depends on the source of the former. All these sources vanish for constancy of equilibrium concentration and kinetic temperatures. Elimination of the sources leads in each case to sixth-order differential equations.

The works of Wait,[1] Budden,[2] Tidman,[16] and Bremmer[20] show that the method of WKB would be suitable for studying wave propagation in slowly varying plasma. Pioneering research on plasma wave coupling was done by Vlasov,[13] Tidman,[16] and Field.[19] Chakraborty[22] also worked on coupling of waves in Vlasov plasma in a crude and simplified way. Propagation of transverse waves in a slowly varying fluid plasma for electron perturbation only was studied by the author.[23]

Section 2 begins with the plasma field equations. The equations of coupling between **H** and density perturbation of both electrons and ions are then established. In Sec. 3 the energy conservation is deduced for the plasma model under consideration. Section 4 considers the case of coupling between low-frequency ion motion and electromagnetic waves. Evaluation of power transfer from ion acoustic to transverse field and vice versa is the subject matter of Sec. 5.

2. THE COUPLING EQUATIONS

Starting linearized equations for the two-component fully ionized plasma in the absence of a static magnetic field can be taken in the form

$$N\frac{\partial \mathbf{U}_e}{\partial t} = -\text{grad}\,(V_e^2\delta N_e) - \frac{eN}{m}\mathbf{E}, \tag{2.1}$$

$$N\frac{\partial \mathbf{U}_i}{\partial t} = -\text{grad}\,(V_i^2\delta N_i) + \frac{eN}{M}\mathbf{E}, \tag{2.2}$$

$$\text{curl}\,\mathbf{E} = -\frac{1}{C}\frac{\partial \mathbf{H}}{\partial t}, \tag{2.3}$$

$$\text{curl}\,\mathbf{H} = \frac{1}{C}\frac{\partial \mathbf{E}}{\partial t} + \frac{4\pi N_e}{C}(-\mathbf{U}_e + \mathbf{U}_i), \tag{2.4}$$

$$\text{div}\,(N\mathbf{U}_e) + \frac{\partial}{\partial t}\delta N_e = 0, \quad \text{div}\,(N\mathbf{U}_i) + \frac{\partial}{\partial t}\delta N_i = 0, \tag{2.5}$$

$$\text{div}\,\mathbf{E} = -4\pi e(\delta N_e - \delta N_i), \tag{2.6}$$

where the quantities bearing the subscripts e and i are those for the electrons and ions respectively, $V_e^2 = \chi T_e/m$ and $V_i^2 = \chi T_i/M$, χ is the Boltzmann constant, and T_e and T_i are the electron and ion kinetic temperatures, respectively. The meanings of the other quantities are obvious and standard. The equilibrium parameters N, V_i (or T_i), and V_e (or T_e) are the slowly varying functions of position.

The volume forces due to nonzero gradients of equilibrium pressure may be maintained by some static nonelectrical forces acting in the same direction on both electrons and ions and sustaining the given plasma configuration without causing charge separation. Hence these forces are independent of charge separation. This approximation can roughly represent some astrophysical models, the ionosphere, or charged cloud causing lightning. Even a slow drift of the body owing to static forces proportional to the gradients of the earlier mentioned parameters can be allowed. For simplicity, theoretical analysis of the balance of equilibrium forces has been dropped because the consequent inclusion of complications cannot throw any additional light on the problems of this paper.

$$\text{In the uniform equilibrium plasma grad } N = 0, \cdots. \tag{2.7}$$

In the plasma model of this paper this relation is not true.

The equations of coupling should link the purely transverse $\mathbf{H}$ and purely longitudinal δN_e and δN_i when (2.7) does not hold. Also, the number of coupling equations should be sufficient.

Equation (2.4) is differentiated with respect to time. From that, $\mathbf{U}_e$ and $\mathbf{U}_i$ are eliminated with the help of momentum transfer equations (2.1) and (2.2). Lastly, $\mathbf{E}$ is eliminated with the help of electrodynamical equations, giving for $\mathbf{H}$ waves the equation

$$\dddot{\mathbf{H}} + (2w_e^2 + 2w_i^2 - C^2\nabla^2)\ddot{\mathbf{H}} + (W_e^2 + W_i^2)(W_e^2 + W_i^2 - C^2\nabla^2)\dot{\mathbf{H}} - C^2[\text{grad}\,(W_e^2 + W_i^2)\,\text{curl}\,\dot{\mathbf{H}}]$$
$$= -4\pi ec\left[\text{grad}\,(W_e^2 + W_i^2)\,\text{grad}\left(\frac{\delta p_e}{m} - \frac{\delta p_i}{M}\right)\right]. \tag{2.8}$$

The equations for δN_e and δN_i are

$$\ddot{\delta N}_e + (2W_e^2 + W_i^2 - \nabla^2 V_e^2)\delta\ddot{N}_e - W_e^2\delta\ddot{N}_i + (W_e^2 + W_i^2)(W_e^2 - \nabla^2 V_e^2)\delta N_e - (W_e^2 + W_i^2)W_e^2\delta N_i$$
$$+ \left(\text{grad}\left(\frac{\delta p_e}{m} - \frac{\delta p_i}{M}\right)\text{grad}\,W_e^2\right) = \frac{C}{4\pi e}(\text{curl}\,\dot{\mathbf{H}}\,\text{grad}\,W_e^2), \tag{2.9}$$

$$\ddot{\delta N}_i + (2W_i^2 + W_e^2 - \nabla^2 V_i^2)\delta\ddot{N}_i - W_i^2\delta\ddot{N}_e + (W_e^2 + W_i^2)(W_i^2 - \nabla^2 V_i^2)\delta N_i - (W_e^2 + W_i^2)W_i^2\delta N_e$$
$$- \left(\text{grad}\left(\frac{\delta p_e}{m} - \frac{\delta p_i}{M}\right)\text{grad}\,W_i^2\right) = -\frac{C}{4\pi e}(\text{curl}\,\dot{\mathbf{H}}\,\text{grad}\,W_i^2), \tag{2.10}$$

where

$$\delta p_e = mV_e^2\delta N_e, \quad \delta p_i = MV_i^2\delta N_i, \quad W_e^2 = 4\pi Ne^2/m, \; W_i^2 = 4\pi Ne^2/M.$$

These have been obtained by eliminating $\mathbf{U}_e$ and $\mathbf{U}_i$ between (2.5), (2.1), and (2.2) and then eliminating $\mathbf{E}$ with the help of (2.3) and (2.4).

The first of these three equations gives the usual dispersion relation for transverse waves and the last two for longitudinal waves if N, V_e, and V_i are uniform and a plane wave solution is imposed. Also, coupling ceases if the disturbance propagates along the direction of gradients of these parameters.

If time dependence is exp $(-iWt)$, then (2.8)–(2.10) yield

$$(W^2 - W_e^2 - W_i^2 + C^2\nabla^2)\mathbf{H} - \frac{C^2[\text{grad}\,(W_e^2 + W_i^2)\,\text{curl}\,\mathbf{H}]}{W^2 - W_e^2 - W_i^2}$$
$$= \frac{4\pi iec[\text{grad}\,(W_e^2 + W_i^2)\,\text{grad}\,[(\delta p_e/m) - (\delta p_i/M)]]}{W(W^2 - W_e^2 - W_i^2)}, \tag{2.11}$$

$$(W^2 - W_e^2 + \nabla^2 V_e^2)\delta N_e + W_e^2\delta N_i + \frac{(\text{grad}\,[(\delta p_e/m) - (\delta p_i/M)]\,\text{grad}\,W_e^2)}{W^2 - W_e^2 - W_i^2} = -\frac{iWC(\text{curl}\,\mathbf{H}\,\text{grad}\,W_e^2)}{4\pi e(W^2 - W_e^2 - W_i^2)}, \tag{2.12}$$

$$(W^2 - W_i^2 + \nabla^2 V_i^2)\delta N_i + W_i^2\delta N_e - \frac{(\text{grad}\,[(\delta p_e/m) - (\delta p_i/M)]\,\text{grad}\,W_i^2)}{W^2 - W_e^2 - W_i^2} = \frac{iWc(\text{curl}\,\mathbf{H}\,\text{grad}\,W_i^2)}{4\pi e(W^2 - W_e^2 - W_i^2)}. \tag{2.13}$$

From these equations sixth-order differential equations for each of δN_e, δN_i, and $\mathbf{H}$ would be obtained by eliminating the other two.

3. ENERGY TRANSPORT

Energy transport due to propagation of waves in uniform plasma in the linear approximation has been thoroughly discussed in Refs. 6, 7, 24, and 25. Following (7) and extending to the nonuniform model of Sec. 2, we see that the power flow density $(\mathbf{E}\cdot\mathbf{J})$ and the energy conservation are given by

$$(\mathbf{E}\cdot\mathbf{J}) = \frac{\partial}{\partial t}(W_K + W_P) + \operatorname{div}\mathbf{S}' + L = -\operatorname{div}\mathbf{S} - \frac{\partial}{\partial t}(W_E + W_M), \tag{3.1}$$

where

$$W_K = \tfrac{1}{2}(MNU_i^2 + mNU_e^2),$$
$$W_P = (\delta p_i \delta N_i + \delta p_e \delta N_e)/2N, \tag{3.2}$$
$$\mathbf{S}' = \mathbf{U}_i\delta p_i + \mathbf{U}_e\delta p_e, \quad \mathbf{S} = (C/4\pi)[\mathbf{EH}],$$
$$W_E = E^2/8\pi, \quad W_M = H^2/8\pi, \tag{3.3}$$
$$L = [\delta p_i(\mathbf{U}_i \operatorname{grad} N) + \delta p_e(\mathbf{U}_e \operatorname{grad} N)]/N. \tag{3.4}$$

The volume source term L vanishes for (2.7).

4. COUPLING BETWEEN ELECTROMAGNETIC AND ION WAVES

Low-frequency disturbances in the plasma can excite propagation of ion oscillations, interaction of which with inhomogeneities gives rise to radiation of electromagnetic waves through the boundary of a nonuniform region. In this case one can put $W_e = 0$ and $\delta N_e = 0$ and consider only equations (2.11) and (2.13) in the form

$$\nabla^2\mathbf{H} + K_t^2\mathbf{H} = \frac{[\operatorname{grad} W_i^2 \operatorname{curl}\mathbf{H}]}{K_t^2C^2} + \frac{4\pi ie[\operatorname{grad} W_i^2 \operatorname{grad}\delta p]}{MWcK_1^2V_i^2}, \tag{4.1}$$

$$\nabla^2\delta p + K_1^2\delta p = -\frac{(\operatorname{grad}\delta p \operatorname{grad} W_i^2)}{K_1^2V_i^2} + \frac{iWc(\operatorname{curl}\mathbf{H} \operatorname{grad} W_i)}{4\pi eMV_i^2K_1^2}, \tag{4.2}$$

where

$$\frac{W^2 - W_i^2}{C^2} = K_t^2, \quad \frac{W^2 - W_i^2}{V_i^2} = K_1^2, \quad \delta p = V_i^2\delta p_i. \tag{4.3}$$

They will be solved in a region free from boundaries and sources of singularities. The density perturbation falls from outside and passes through the region exciting modes of $\mathbf{H}$. Gradients of N and T_i are assumed to be acting along (O, Z). The disturbance propagates in the (Z, X) plane and the x dependence of the field variables is $\exp(iK_0x)$, where K_0 is constant. Then (4.1), resolved into the component equations, and (4.2) yield

$$\frac{d^2Hx}{dZ^2} + \frac{1}{W^2 - W_i^2}\frac{dW_i^2}{dZ}\frac{dHx}{dZ} + (K_t^2 - K_0^2)Hx - \frac{iK_0}{W^2 - W_i^2}\frac{dW_i^2}{dz}Hz = 0, \tag{4.4}$$

$$\frac{d^2}{dZ^2}Hz + (K_t^2 - K_0^2)Hz = 0, \tag{4.5}$$

$$\frac{d^2}{dZ^2}Hy + \frac{1}{W^2 - W_i^2}\frac{dW_i^2}{dZ}\frac{dHy}{dZ} + (K_t^2 - K_0^2)Hy = -\frac{4\pi eK_0}{MWcK_1^2}\frac{dW_i^2}{dZ}\delta p_i, \tag{4.6}$$

$$\frac{d^2}{dZ^2}\delta p - \frac{d}{dZ}\log(W^2 - W_i^2)\frac{d}{dZ}\delta p + (K_1^2 - K_0^2)\delta p = -\frac{K_0WcM}{4\pi e(W^2 - W_i^2)}\frac{dW_i^2}{dZ}Hy. \tag{4.7}$$

Of these four equations the first and second connect together Hx and Hz (the components of $\mathbf{H}$ lying in the plane of incidence), which are not coupled to δp by grad N. But δp is coupled by grad N to Hy through (4.6) and (4.7)—equations useful for our purpose and reducible to

$$\left(\frac{d^2}{dZ^2} + K_1^2\right)\psi_1 = \epsilon\psi_2, \quad \left(\frac{d^2}{dZ^2} + K_2^2\right)\psi_2 = \epsilon\psi_1, \tag{4.8}$$

where

$$K_1^2 = K_t^2 - K_0^2 - K_t\frac{d^2}{dZ^2}\left(\frac{1}{K_t}\right),$$
$$K_2^2 = K_1^2 - K_0^2 - K_t\frac{d^2}{dZ^2}\left(\frac{1}{K_t}\right), \tag{4.10}$$

$$\epsilon = \frac{d}{dZ}(K_0\log K_t^2) = -\frac{K_0}{W^2 - W_0^2}\frac{dW_i^2}{dZ}, \tag{4.11}$$

$$\psi_1 = \frac{Hy}{K_t}, \quad \psi_2 = \frac{4\pi e\delta p}{MWcK_t}. \tag{4.12}$$

For $\epsilon = 0$, the WKB solutions are

$$\psi_1 = \frac{A_1\exp\left(i\int K_1\,dZ\right) + B_1\exp\left(-i\int K_1\,dZ\right)}{\sqrt{K_1}},$$
$$\psi_2 = \frac{A_2\exp\left(i\int K_2\,dZ\right) + B_2\exp\left(-i\int K_2\,dZ\right)}{\sqrt{K_2}}. \tag{4.13}$$

Then the complete solution, terms being kept up to derivatives of A's and B's, would be

$$\psi_1 = \frac{Hy}{K_t} = \frac{A_1 e^{i\xi_1} + B_1 e^{-i\xi_1}}{\sqrt{K_1}} + \frac{K_2^2}{\epsilon}\left[\frac{A_2 e^{i\xi_2}}{\sqrt{K_2}}\left(-\frac{1}{\sqrt{K_2}}\frac{d^2(\sqrt{K_2})}{d\xi_2^2} + 2\frac{id}{d\xi_2^2}\log A_2\right) + \frac{B_2 e^{-i\xi_2}}{\sqrt{K_2}}\left(-\frac{1}{\sqrt{K_2}}\frac{d^2(\sqrt{K_2})}{d\xi_2^2} - 2\frac{id}{d\xi_2^2}(\log B_2)\right)\right], \tag{4.14}$$

$$\delta p = \frac{MWcKt}{4\pi e}\left[\frac{A_2 e^{i\xi_2} + B_2 e^{-i\xi_2}}{\sqrt{K_2}} + \frac{K_1^2}{\epsilon}\frac{A_1 e^{i\xi_1}}{\sqrt{K_1}} \times \left(-\frac{1}{\sqrt{K_1}}\frac{d^2}{d\xi_1^2}(\sqrt{K_1}) + 2i\frac{d}{d\xi_1}(\log A_1)\right) + \frac{K_1^2}{\epsilon}\frac{B_1 e^{-i\xi_1}}{\sqrt{K_1}} \times \left(-\frac{1}{\sqrt{K_1}}\frac{d^2}{d\xi_1^2}(\sqrt{K_1}) - 2i\frac{d}{d\xi_1}\log B_1\right)\right], \tag{4.15}$$

where

$$\xi_1 = \int K_1\, dZ, \quad \xi_2 = \int K_2\, dZ. \tag{4.16}$$

Putting the first or second solution of ψ_1 from (4.13) on the fourth-order equation for ψ, obtained by eliminating ψ_2 between the equations of (4.8), we get

$$\frac{d}{d\xi_1}\log(A_1, B_1) = \frac{i}{2}\left(\frac{\epsilon^2}{K_1^2(K_2^2 - K_1^2)} + \frac{1}{\sqrt{K_1}}\frac{d^2(\sqrt{K_1})}{d\xi_1^2}\right)(-1, 1), \tag{4.17}$$

where higher-order derivatives of A_1 and B_1 have been neglected. Similarly, the values of derivatives of $\log A_2$ and $\log B_2$ are calculated by interchanging the suffixes 1 and 2 in (4.17). In this way the variation in amplitude is expressed in terms of variation of the material parameters.

Hence (4.14) and (4.15) give

$$\frac{Hy}{K_t} = \frac{A_1 e^{i\xi_1} + B_1 e^{-i\xi_1}}{\sqrt{K_1}} + \frac{K_2^2}{2\epsilon}\left(\frac{A_2 e^{i\xi_2} + B_2 e^{-i\xi_2}}{\sqrt{K_2}}\right) \times \left(\frac{\epsilon^2}{K_2^2(K_1^2 - K_2^2)} - \frac{1}{\sqrt{K_2}}\frac{d^2}{d\xi_2^2}(\sqrt{K_2})\right), \tag{4.18}$$

$$\delta p\left(\frac{4\pi e}{MWcK_t}\right) = \frac{A_2 e^{i\xi_2} + B_2 e^{-i\xi_2}}{\sqrt{K_2}} + \frac{K_1^2}{2\epsilon}\left(\frac{A_1 e^{i\xi_1} + B_1 e^{-i\xi_1}}{\sqrt{K_1}}\right) \times \left(\frac{\epsilon^2}{K_1^2(K_2^2 - K_1^2)} - \frac{1}{\sqrt{K_1}}\frac{d^2}{d\xi_1^2}(\sqrt{K_1})\right). \tag{4.19}$$

The coefficients of $e^{i\xi_2}$ and $e^{-i\xi_2}$ in (4.18) and of $e^{i\xi_1}$ and $e^{-i\xi_1}$ in (4.19) are small in the sense of smallness of quantities adopted in this paper.

5. POWER TRANSFER FROM ION ACOUSTIC TO TRANSVERSE FIELDS

The direction normal to the surface

$$K_0 x + K_1 \int dZ = \text{const}$$

is regarded as the incident direction, and the corresponding wave is called the incident wave. Similarly, the normal to the surface

$$K_0 x - \int K\, dZ = \text{const}$$

is the direction of reflection. Also the suffixes i_1 and r_1 would be used to denote the incident and the reflected field components for the wave whose $K = K_1$ of (4.10). Similar quantities for $K = K_2$ of (4.10) would bear the suffixes i_2 and r_2.

In the uniform equilbrium plasma i_1 and r_1 fields (respectively, i_2 and r_2) are the forward and backward going parts of the transverse (respectively, longitudinal) waves. Hence, as continuation in the slowly varying equilibrium plasma though the waves are mutually coupled, the i_1 and r_1 fields are regarded as transverse and the i_2 and r_2 as longitudinal.

The Z components of $\mathbf{S}$ and $\mathbf{S}'$ of (3.3) become

$$S_z = -\frac{iWHy}{4\pi K_t^2}\left(\frac{dHy}{dZ} + \frac{4\pi e K_0}{MWc}\delta p\right), \tag{5.1}$$

$$S_z' = -\frac{i(\delta p)}{M(W^2 - W_1^2)}\left(\frac{W}{N}\frac{d}{dZ}(\delta p) + K_0 ceHy\right). \tag{5.2}$$

The Poynting vector for the i_1 and r_1 fields has a principal term which does not vanish and a small term which vanishes for constancy of N. In that case, for i_2 and r_2 fields also, $\mathbf{S}$ vanishes, but, if N is slowly varying, then $\mathbf{S}$ is small.

Keeping the principal terms and the largest of the smaller terms, calculating the real values of S_z and S_z' in all the incident cases, and then taking the average

over one time period, one gets

$$\langle S_z\rangle^{i_1} = -\frac{A_1^2}{K_1}\frac{K_1 W}{8\pi}\left[1+\frac{1}{2}\left(\frac{\epsilon^2}{K_1^2(K_2^2-K_1^2)} + \frac{1}{\sqrt{K_1}}\frac{d^2}{d\xi_1^2}(\sqrt{K_1})\right)\right], \tag{5.3}$$

$$\langle S_z\rangle^{i_2} = -\frac{A_2^2}{K_2}\frac{K_2 W\epsilon^2}{8\pi(K_1^2-K_2^2)}, \tag{5.4}$$

$$\langle S_z'\rangle^{i_1} = -\frac{A_1^2}{K_1}\frac{K_1 W^3\epsilon^2}{8\pi W_i^2(K_2^2-K_1^2)}, \tag{5.5}$$

$$\langle S_z'\rangle^{i_2} = -\frac{A_2^2}{K_2}\frac{K_2 W^3}{8\pi W_i^2}\left[1+\frac{1}{2}\left(\frac{\epsilon^2}{K_2^2(K_1^2-K_2^2)} + \frac{1}{\sqrt{K_2}}\frac{d^2}{d\xi_2^2}(\sqrt{K_2})\right)\right]. \tag{5.6}$$

The ratios

$$\frac{\langle S_z'\rangle^{i_1}}{\langle S_z\rangle^{i_1}} = \frac{K_0^2 W^2 W_i^2 V_i^4}{(W^2-W_i^2)^4}\left(\frac{d}{dz}\log W_i^2\right)^2,$$

$$\frac{\langle S_z\rangle^{i_2}}{\langle S_z'\rangle^{i_2}} = \frac{K_0^2 V_i^4 W_i^6}{W^2(W^2-W_i^2)^4}\left(\frac{d}{dz}\log W_i^2\right)^2 \tag{5.7}$$

can be regarded as the coupling coefficients up to first order of derivatives of N.

The amplitudes proportional to A_1, A_2, etc., in (5.3)–(5.6) can be replaced by

$$\frac{A_1}{\sqrt{K_1}} = (Hy\, Hy^*)^{\frac{1}{2}}, \quad \frac{A_2}{\sqrt{K_2}} = (\delta p\, \delta p^*)^{\frac{1}{2}}\left(\frac{4\pi e}{MWcK_t}\right),$$

where Hy^* and δp^* are complex conjugates of Hy and δp.

We now consider slowly varying plasma between two parallel planes $Z = Z_1$ and $Z = Z_2$. Electromagnetic waves fall obliquely on the side $Z = Z_1$ and pass out through the plane $Z = Z_2$. Interaction with the plasma generates, at all points of its passage pressure wave, the net power flow, the form of which is $[\langle S_z'\rangle^{i_1}]_1^2$, where the suffixes 1 and 2 on the right outside the square brackets mean the difference between the values at the two end planes. Hence

$$[\langle S_z'\rangle^{i_1}]_1^2 = \left[\frac{|Hy|^2 W_i^2 V_i^4(1-W_i^2+K_0^2C^2/W^2)^{\frac{1}{2}}}{(1-W_i^2/W^2)^5} \times\left(\frac{d}{dz}\log W_i^2\right)^2\right]^2\frac{K_0^2 C}{8\pi W^6}.$$

The quantity $|Hy|$ is the magnitude of the amplitude of the magnetic field. Similarly,

$$[\langle S_z\rangle^{i_2}]_1^2 = \frac{K_0^2}{2W^{10}}\left[\frac{|\delta p|^2 W_i^6 V_i^3(1-W_i^2+K_0^2c^2/W^2)^{\frac{1}{2}}}{P_i(1-W_i^2/W^2)^5} \times\left(\frac{d}{dz}\log W_i^2\right)^2\right]_1^2, \quad (P_i = MN),$$

is the Poynting flow created by pressure wave falling obliquely on $Z = Z_1$ and passing out through $Z = Z_2$.

[1] J. R. Wait, *Electromagnetic Waves in Stratified Media* (Pergamon, New York, 1962).
[2] K. G. Budden, *Radio Waves in Ionosphere* (Cambridge U.P., New York, 1961).
[3] T. H. Stix, *The Theory of Plasma Waves* (McGraw-Hill, New York, 1962).
[4] L. M. Brekhovskikh, *Waves in Layered Media* (Academic, New York, 1960).
[5] V. L. Ginzburg, *Propagation of Electromagnetic Waves in Plasmas* (Gordon and Breach, New York, 1961).
[6] W. P. Allis, S. J. Buchsbaum, and A. Bers, *Waves in Anisotropic Plasmas* (M.I.T. Press, Cambridge, Mass., 1963).
[7] J. F. Denisse and J. L. Delroix, *Plasma Waves* (Interscience, New York, 1963).
[8] H. Bremmer, *Terrestrial Radio Waves* (Elsevier, Amsterdam, New York, 1949).
[9] H. Bremmer, "Propagation of Electromagnetic Waves" in *Handbuch der Physik, Vol. 16*, S. Flügge, Ed. (Springer, Berlin, 1958), p. 423.
[10] *Theory of Electromagnetic Waves: A Symposium, 1951*, S. A. Schelkunoff, Ed. (Interscience, New York, 1951).
[11] N. Bloembergen, *Nonlinear Optics* (Benjamin, New York, 1965).
[12] S. A. Akhmanov and R. V. Khokhlov, *Problemi Nelineinei Optiki* (in Russian) (Moscow, 1962).
[13] G. Bekefi, *Radiation Processes in Plasmas* (Wiley, New York, 1966).
[14] J. Dawson, "Radiation from Plasmas," in *Advances in Plasma Physics*, A. Simon and W. B. Thompson, Eds. (Interscience, New York, 1968), Vol. I, Chap. I.
[15] A. A. Vlasov, Oochionie Zap., M.G.Y., **75**, No. 2 (1945).
[16] D. A. Tidman, Phys. Rev. **117**, 366 (1960).
[17] A. H. Kritz and D. Mintzer, Phys. Rev. **117**, 382 (1960).
[18] K. H. Dunphy, D. Kahn, and D. Mintzer, Phys. Fluids **10**, 162 (1967).
[19] G. B. Field, Astrophys. J. **124**, 555 (1956).
[20] H. Bremmer, Physica **15**, 593 (1949).
[21] (a) L. A. Chernov, *Wave Propagation in a Random Medium* (McGraw-Hill, New York, 1960). (b) V. I. Tatarskii, *Wave Propagation in a Turbulent Medium* (McGraw-Hill, New York, 1960).
[22] B. Chakraborty, thesis, Physics Faculty, Moscow State University, 1967 (in Russian), Chap. 3.
[23] B. Chakraborty, J. Math. Phys. **11**, 2570 (1970).
[24] K. A. Haynes and D. Kahn, Phys. Fluids **8**, 1681 (1965).
[25] M. H. Cohen, Phys. Rev. **126**, 389 (1962).

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 12, NUMBER 3 MARCH 1971

Some Remarks on the Theory of Trapping of Excitons in the Photosynthetic Unit

J. W. SANDERS, TH. W. RUIJGROK, AND J. J. TEN BOSCH
Rijksuniversiteit, Utrecht, The Netherlands

(Received 20 April 1970)

The problem of random motion of an exciton in a square lattice with a sink is formulated in terms of finite difference equations. The case where the exciton suffers a fixed delay time at each lattice site is solved numerically. The more realistic case, also investigated by Knox [J. Theoret. Biol. **21**, 244 (1968)], where this delay time is inversely proportional to the number of nearest neighbors, is shown to be equivalent to the case with periodic boundary conditions. Consequently, an exact analytic expression for the total delay time can for that case be derived. For lattices of any size this quantity can now easily be calculated. Giving the exciton the possibility of escaping from the lattice will influence the total delay time. This effect is investigated, as well as the effect of distributing the sink over the lattice sites. As a by-product, an amusingly simple formula

$$k(\lambda) = \frac{1}{c(\lambda)} \sum_{\mu} c(\mu)$$

is derived for the average path length upon first return in a random walk over an arbitrary network.

1. INTRODUCTION

The subject of trapping of energy by a sink in a plane lattice of molecules, in which a localized excitation can move by random jumps to neighboring molecules and finally to a trap, has been treated by several authors: Avery, Bay, and Szent-Györgyi,[1] ten Bosch and Ruijgrok,[2] Pearlstein,[3] Robinson,[4] Knox,[5] and Montroll.[6,7] About the timing of the jumps, two different assumptions can be made:

(1) The time spent on a site is inversely proportional to, or otherwise dependent upon, the number of neighbors of the site.

(2) The time spent on a site is fixed.

The first case occurs, for instance, when the jump is caused by the interaction of a populated site with its nearest (unpopulated) neighbors. This seems to be the case in the photosynthetic unit (e.g., see Duysens[8]). The second case would apply if the jump were caused either by an internal clock mechanism or by the interaction with all other lattice sites. In both cases difference equations have been derived[2,5] for the average time $T(x, y)$ between the moments of creation of the exciton in the lattice point (x, y) and of its annihilation by a sink in the origin (0, 0).

The first purpose of this paper is to give a unified derivation of the equations for the two cases and to show how in these two cases the average life time of an exciton depends on the size of the lattice (Sec. 2). For case (1) we will derive an exact analytic expression [Eq. (4.2)] for $T(x, y)$, which can be used for its numerical calculation for a square lattice of any size. For case (2) numerical calculations are performed, and the results are compared with those of case (1).

In Sec. 3 we investigate to what extent this average life time is changed when the exciton has the possibility of escaping from the lattice before it is trapped, e.g., by a fluorescence mechanism.

It is frequently assumed that imposing periodic boundary conditions introduces only minor deviations. For square lattices of different size, these deviations are calculated in Sec. 4. It turns out that such deviations only occur when the exciton spends a fixed time on every lattice site it visits, i.e., in case (2). In Sec. 5 a study is made of the case where the trapping center can be found on any site with an arbitrary distribution function.

2. THE AVERAGE TRAPPING TIME

Consider the problem of a random walk of an exciton on a 2-dimensional square lattice with $(2n + 1)^2$ lattice points. The lattice distance is taken to be unity. The exciton jumps with equal probability to any of its nearest neighbors. It disappears if it arrives at the center (0, 0), where the trap is located. The center of the lattice is taken as the origin of the coordinate system, so that a lattice site has the coordinates (x, y), x and y being integers with $|x| \leq n$ and $|y| \leq n$. Three types of points are distinguished (see Fig. 1): an internal point, e.g., A has four nearest neighbors; a side point, e.g., B has three nearest neighbors; a corner point, e.g., C has two nearest neighbors.

In case (1) an exciton, before jumping from the point P to one of its nearest neighbors, will spend a time τ on P, if P is an internal point. This time becomes $\frac{4}{3}\tau$ if P is a side point and 2τ if P is a corner point. In general, a time $4\tau/c(x, y)$ is spent on the

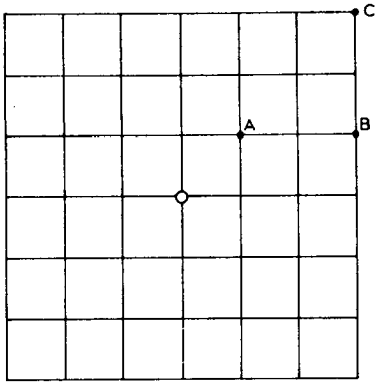

FIG. 1. Three types of points are distinguished.

lattice site (x, y), where $c(x, y)$ is the number of nearest neighbors of (x, y).

Let $T^{(1)}(x, y)$ be the average trapping time of an exciton, which starts at (x, y). For an internal point, different from the origin, one has

$$T^{(1)}(x, y) = \tau + \tfrac{1}{4}[T^{(1)}(x+1, y) + T^{(1)}(x-1, y) + T^{(1)}(x, y+1) + T^{(1)}(x, y-1)]. \quad (2.1)$$

This equation is also correct if (x, y) is a nearest neighbor of the origin. The $T^{(1)}(0, 0)$ occurring in the right-hand side of (2.1) must then be put equal to zero. If (x, y) is a side point (e.g., B in Fig. 1), the equation becomes

$$T^{(1)}(x, y) = \tfrac{4}{3}\tau + \tfrac{1}{3}[T^{(1)}(x-1, y) + T^{(1)}(x, y+1) + T^{(1)}(x, y-1)]. \quad (2.2)$$

For a corner point (e.g., C in Fig. 1), one gets

$$T^{(1)}(x, y) = 2\tau + \tfrac{1}{2}[T^{(1)}(x-1, y) + T^{(1)}(x, y-1)]. \quad (2.3)$$

The equations for $T^{(1)}(x, y)$ can be combined into

$$T^{(1)}(x, y) = \frac{4\tau}{c(x, y)} + \frac{1}{c(x, y)} \sum_{x', y'}^{(x, y)} T^{(1)}(x', y'), \qquad (x, y) \neq (0, 0),$$
$$T^{(1)}(0, 0) = 0. \quad (2.4)$$

The summation extends over the nearest neighbors of (x, y). The set of equations (2.4) is identical with Eq. (21) of the paper by Knox,[5] if one identifies F^{-1} with 4τ. For reasons of simplicity we consider, instead of $T^{(1)}(x, y)$, the quantity

$$D^{(1)}(x, y) = \tau^{-1} T^{(1)}(x, y).$$

For this, the following equations hold:

$$D^{(1)}(x, y) = \frac{4}{c(x, y)} + \frac{1}{c(x, y)} \sum_{x', y'}^{(x, y)} D^{(1)}(x', y'), \qquad (x, y) = (0, 0),$$
$$D^{(1)}(0, 0) = 0. \quad (2.5)$$

It should be stressed that for this case $D^{(1)}(x, y)$ must not be considered as an average path-length.

In case (2) the exciton spends a fixed time τ on each lattice site. The equations for the average trapping time $T^{(2)}(x, y)$ differ from (2.4) only in the first term on the right-hand side. This should read τ. The corresponding equations for

$$D^{(2)}(x, y) = \tau^{-1} T^{(2)}(x, y)$$

become

$$D^{(2)}(x, y) = 1 + \frac{1}{c(x, y)} \sum_{x', y'}^{(x, y)} D^{(2)}(x', y'), \qquad (x, y) \neq (0, 0),$$
$$D^{(2)}(0, 0) = 0, \quad (2.6)$$

which are the equations derived before by ten Bosch and Ruijgrok.[2]

In this way it is seen that the difference between the two sets of equations (2.5) and (2.6) can be traced back to the assumptions (1) and (2) of Sec. 1. One should realize that in both cases the average path length is the same and is given by $D^{(2)}(x, y)$. As has been pointed out by Knox,[5] the earlier solution[2] of (2.6) was at fault.

For several values of n the functions $D^{(1)}(x, y)$ and $D^{(2)}(x, y)$ were numerically solved from their respective equations. As a check on the computation of $D^{(2)}(x, y)$, use was made of the formula

$$D^{(2)}(1, 0) = 4n^2 + 2n - 1, \quad (2.7)$$

which is proved in Sec. 5. It is the extension of a formula, derived by Montroll,[6] for a lattice with periodic boundary conditions.

For $n = 9$ the functions $D^{(1)}(x, y)$ and $D^{(2)}(x, y)$ are given in Tables I and II, respectively. The averages over all lattice points of $D^{(1)}(x, y)$ and $D^{(2)}(x, y)$, i.e.,

$$D_1(n) = \frac{1}{(2n+1)^2} \sum_{x, y} D^{(1)}(x, y)$$

TABLE I. The average trapping time $D^{(1)}(x, y)$ (in units of τ) for a square lattice with $n = 9$ in the case of a variable delay time [case (1)].

y										
9	786	787	791	796	802	809	816	821	825	827
8	779	781	785	791	798	806	813	818	823	
7	766	768	773	781	789	799	807	814		
6	745	748	755	765	777	788	799			
5	715	719	730	744	760	775				
4	672	679	696	718	740					
3	612	626	655	687						
2	521	553	605							
1	360	458								
0	0									
	0	1	2	3	4	5	6	7	8	9 x

TABLE II. The average trapping time $D^{(2)}(x, y)$ (in units of τ) for a square lattice with $n = 9$ in the case of a fixed delay time [case (2)].

y										
9	740	741	745	749	755	761	767	771	774	775
8	735	736	740	745	752	759	765	769	773	
7	723	725	730	736	745	753	760	766		
6	704	706	713	722	733	743	753			
5	676	680	690	703	718	732				
4	636	643	659	679	699					
3	580	592	619	650						
2	493	524	573							
1	341	433								
0	0									
	0	1	2	3	4	5	6	7	8	9 x

and

$$D_2(n) = \frac{1}{(2n+1)^2} \sum_{x,y} D^{(2)}(x, y),$$

have been calculated and are plotted in Fig. 2.

After the numerical calculation of $D^{(1)}(x, y)$ [Eq. (2.5)] was carried out, it was observed that the result was identical to the solution for a system with periodic boundary conditions, which is treated in detail in Sec. 4. We now prove the above-mentioned identity by showing that $D^{(1)}(x, y)$, defined as the solution of (2.5), also satisfies Eq. (4.1) for the periodic case.

If $(x, y) \neq (0, 0)$ is a point of the type A (see Fig. 1), Eq. (2.5) becomes

$$D^{(1)}(x, y) = 1 + \tfrac{1}{4}[D^{(1)}(x+1, y) + D^{(1)}(x-1, y) + D^{(1)}(x, y+1) + D^{(1)}(x, y-1)]. \quad (2.8)$$

If (x, y) is a point of the type B, Eq. (2.5) becomes

$$D^{(1)}(n, y) = \tfrac{4}{3} + \tfrac{1}{3}[D^{(1)}(n-1, y) + D^{(1)}(n, y+1) + D^{(1)}(n, y-1)]. \quad (2.9)$$

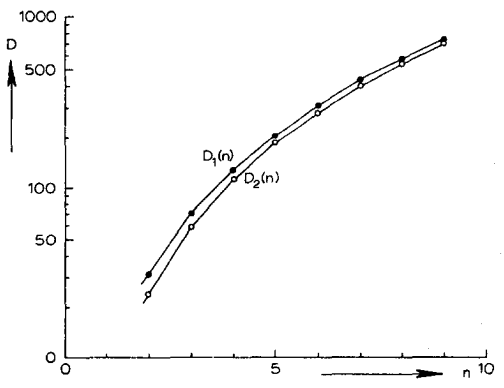

FIG. 2. The average trapping time (in units of τ) for the two cases [(1) variable delay time and (2) fixed delay time] as a function of the size of the lattice.

Since $D^{(1)}(n, y) = D^{(1)}(-n, y)$, we can write (2.9) as

$$D^{(1)}(n, y) = 1 + \tfrac{1}{4}[D^{(1)}(-n, y) + D^{(1)}(n-1, y) + D^{(1)}(n, y+1) + D^{(1)}(n, y-1)] + \tfrac{1}{3} + \tfrac{1}{12}[D^{(1)}(n-1, y) + D^{(1)}(n, y+1) + D^{(1)}(n, y-1)] - \tfrac{1}{4}D^{(1)}(n, y). \quad (2.10)$$

Because of (2.9) the last three terms of (2.10) add up to zero, so that, for a B point, Eq. (2.8) is also satisfied, provided that we read $D^{(1)}(n+1, y)$ as $D^{(1)}(-n, y)$. The same treatment can be given for points on other sides. If (x, y) is a corner point (like C in Fig. 1), Eq. (2.5) becomes

$$D^{(1)}(n, n) = 2 + D(n-1, n), \quad (2.11)$$

which can also be written as

$$D^{(1)}(n, n) = 1 + \tfrac{1}{4}[D^{(1)}(-n, n) + D^{(1)}(n-1, n) + D^{(1)}(n, -n) + D^{(1)}(n, n-1)] + 1 - \tfrac{1}{2}D^{(1)}(n, n) + \tfrac{1}{2}D^{(1)}(n, n-1).$$

Because of (2.11) the last three terms again add up to zero. This completes the proof of the statement that (2.8) and, therefore, (4.1) is satisfied for all $(x, y) \neq (0, 0)$.

In Sec. 4 we give the exact solution of the equations for $D^{(1)}(x, y)$.

As a corollary we derive from (4.2) the fact that $D^{(1)}(1, 0) = D^{(3)}(1, 0) = N - 1$, where N is the total number of lattice points. This is again Montroll's formula.[6]

The most important conclusion from this section is that for a finite lattice, without requiring periodic boundary conditions, an exact and easily calculable analytic expression for the average trapping time can be given.

3. ESCAPE OF THE EXCITON

Consider case (2), in which an exciton spends a fixed time τ on a lattice site, before it jumps. This case will now be modified by assuming that the exciton may decay spontaneously and that its lifetime is τ_0. The probability that in the time τ such a decay has occurred is equal to $\alpha = 1 - e^{-\tau/\tau_0}$. Thus α is supposed to be independent of the number of nearest neighbors. For $\tau \ll \tau_0$ this becomes $\alpha \simeq \tau/\tau_0 \ll 1$.

Let $T_\alpha(x, y)$ be the average time elapsed between the deposition of the exciton at the point (x, y) and either the moment the exciton falls into the trap or the beginning of the interval in which the exciton will

TABLE III. The average total delay time $D_\alpha(x, y)$ (in units of τ) for a square lattice with $n = 9$ and $\alpha = 0.002$ in the case of a fixed delay time [case (2)].

y										
9	304	304	306	307	309	312	314	315	316	317
8	302	302	304	306	308	311	313	315	316	
7	297	298	300	302	305	308	311	313		
6	290	291	294	297	301	305	308			
5	279	281	284	290	295	301				
4	263	266	272	280	288					
3	241	246	257	269						
2	206	218	238							
1	143	181								
0	0									
	0	1	2	3	4	5	6	7	8	9 x

decay. Then $T_\alpha(x, y)$ satisfies

$$T_\alpha(x, y) = (1 - \alpha)\left(\tau + \frac{1}{c(x, y)} \sum_{x', y'}^{(x,y)} T_\alpha(x', y')\right), \quad (x, y) \neq (0, 0),$$

$$T_\alpha(0, 0) = 0 \quad (3.1)$$

or

$$D_\alpha(x, y) = (1 - \alpha)\left(1 + \frac{1}{c(x, y)} \sum_{x', y'}^{(x,y)} D_\alpha(x', y')\right), \quad (x, y) \neq (0, 0),$$

$$D_\alpha(0, 0) = 0, \quad (3.2)$$

with $T_\alpha(x, y) = \tau D_\alpha(x, y)$.

If no trap is present, the time T_α is the same for all lattice points. From (3.1) one then obtains $T_\alpha = (1 - \alpha)\tau/\alpha = (1 - \alpha)\tau_0$, as it should be, regarding

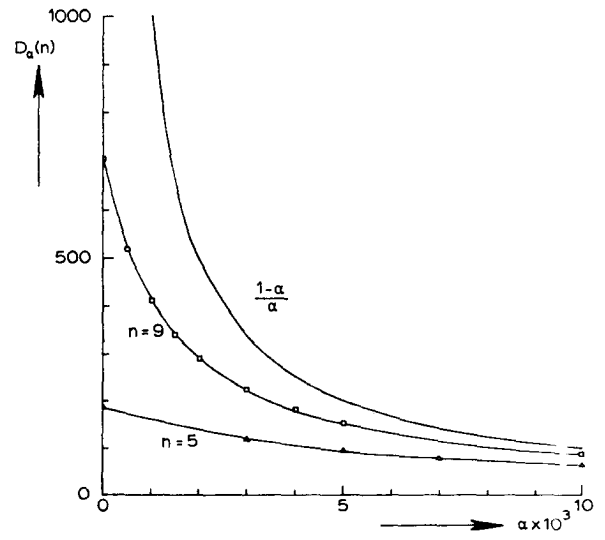

FIG. 3. The average time (in units of τ) between injection and either trapping or escape of the exciton as a function of the probability α that it decays spontaneously in the time τ.

TABLE IV. The average probability of escape $E_\alpha(x, y)$ for a square lattice with $n = 9$ and $\alpha = 0.002$.

y										
9	0.610	0.611	0.613	0.617	0.621	0.625	0.629	0.632	0.634	0.635
8	0.606	0.607	0.610	0.614	0.618	0.623	0.628	0.631	0.634	
7	0.597	0.598	0.602	0.607	0.613	0.619	0.624	0.629		
6	0.582	0.584	0.589	0.596	0.604	0.612	0.619			
5	0.560	0.563	0.571	0.581	0.592	0.603				
4	0.528	0.534	0.546	0.562	0.578					
3	0.483	0.493	0.515	0.540						
2	0.412	0.438	0.477							
1	0.286	0.363								
0	0									
	0	1	2	3	4	5	6	7	8	9 x

the definition of $T_\alpha(x, y)$. The average path length is equal to $D_\alpha(x, y)$. Equations (3.2) have been solved for a number of values of α and n. In Table III we give the result for $\alpha = 0.002$ and $n = 9$. The average value is

$$D_\alpha(n) = \frac{1}{(2n + 1)^2} \sum_{x,y} D_\alpha(x, y), \quad (3.3)$$

which is proportional to the average time the exciton spends on the lattice; it is plotted in Fig. 3 and compared with $(1 - \alpha)/\alpha$ (no trapping center).

Another quantity of interest is the probability that an exciton, which is inserted in the point (x, y) of the lattice, will not be trapped, but will eventually escape. Let this probability be denoted by $E_\alpha(x, y)$. It clearly satisfies the equations

$$E_\alpha(x, y) = \alpha + \frac{(1 - \alpha)}{c(x, y)} \sum_{x', y'}^{(x,y)} E_\alpha(x', y'), \quad (x, y) \neq (0, 0),$$

$$E_\alpha(0, 0) = 0. \quad (3.4)$$

If no trap is present, the escape probability should be unity, which is indeed the solution of (3.4) for this

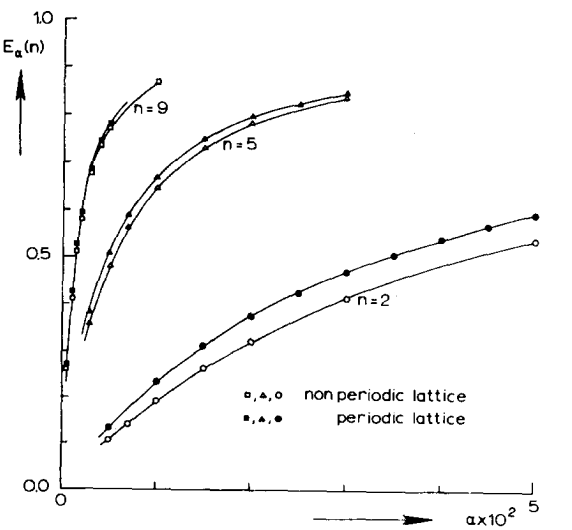

FIG. 4. The probability of escape as a function of the probability α that spontaneous decay of the exciton occurs in the time τ.

case. For $\alpha = 0.002$ and $n = 9$ the values of $E_\alpha(x, y)$, as obtained from (3.4), are given in Table IV. The average escape probability

$$E_\alpha(n) = \frac{1}{(2n+1)^2} \sum_{x,y} E_\alpha(x, y) \tag{3.5}$$

is plotted in Fig. 4 (lower curves for $n = 2$, 5, and 9).

4. PERIODIC BOUNDARY CONDITIONS

In this section an exciton again performs a random walk on a square lattice with $(2n + 1)^2$ sites, one of which is a trap. Case (2) will be supposed to hold, which means that the exciton spends a fixed time τ on a lattice site before it jumps to a nearest neighbor. Periodic boundary conditions hold if this lattice is wrapped on a torus. This means that the point (x, y) is to be identified with the point $(x + 2n + 1, y)$ and also with the point $(x, y + 2n + 1)$. Let the trap be in $(0, 0)$, and let the average time before trapping of an exciton, which starts at (x, y), be indicated by $T^{(3)}(x, y)$. Defining $D^{(3)}(x, y)$ by $T^{(3)}(x, y) = \tau D^{(3)}(x, y)$, we easily see that this function satisfies the equations

$$D^{(3)}(x, y) = 1 + \tfrac{1}{4}[D^{(3)}(x + 1, y) + D^{(3)}(x - 1, y) + D^{(3)}(x, y + 1) + D^{(3)}(x, y - 1)], \quad (x, y) \neq (0, 0)$$

$$D^{(3)}(0, 0) = 0. \tag{4.1}$$

These equations are solved with Fourier transformation. One finds, as can be checked by substitution,

$$D^{(3)}(x, y) = 2 \sum_{j_1 j_2}{}' \frac{\sin^2\{[\pi/(2n+1)](j_1 x + j_2 y)\}}{\sin^2[\pi j_1/(2n+1)] + \sin^2[\pi j_2/(2n+1)]}, \tag{4.2}$$

where $j_1, j_2 = 0, \pm 1, \pm 2, \cdots, \pm n$, but the point $(j_1, j_2) = (0, 0)$ is excluded. We found numerically and proved in Sec. 2 that this function is identical to $D^{(1)}(x, y)$, which was given in Table I.

The average path length

$$D_3(n) = \frac{1}{(2n+1)^2} \sum_{x,y} D^{(3)}(x, y)$$

can also be calculated analytically, and one obtains

$$D_3(n) = \sum_{j_1 j_2}{}' \left[\sin^2\left(\frac{\pi j_1}{2n+1}\right) + \sin^2\left(\frac{\pi j_2}{2n+1}\right) \right]^{-1}. \tag{4.3}$$

In order to find the behavior of this function for large n, one introduces the new variables $x = \pi j_1/(2n + 1)$ and $y = \pi j_2/(2n + 1)$ and the increments $\Delta x = \Delta y = \pi/(2n + 1)$. Equation (4.3) can then be written as

$$D_3(n) = \frac{N}{\pi^2} \sum_{x,y}{}' \frac{\Delta x \Delta y}{\sin^2 x + \sin^2 y}, \tag{4.4}$$

where $N = (2n + 1)^2$ is the total number of lattice sites. For $N \to \infty$ the summation can be replaced by an integration

$$D_3(n) \simeq \frac{N}{\pi^2} \int{}' \frac{dx\, dy}{\sin^2 x + \sin^2 y}. \tag{4.5}$$

If the prime were discarded, the integral would diverge in the origin. The main contribution to $D_3(n)$ can therefore be calculated by excluding a circular region of radius $R = \pi/(2n + 1)$ around the origin. This leads (by introducing polar coordinates) to

$$D_3(n) \simeq \frac{N}{\pi^2} \int_R \int_0^{2\pi} \frac{r\, dr\, d\varphi}{r^2} \quad \text{or} \quad D_3(n) \simeq \frac{N}{\pi} \log N. \tag{4.6}$$

This is the leading term in the following asymptotic formula, derived by Montroll[7]:

$$D_3(n) \simeq (N/\pi) \log N + 0.195056N - 0.1170 - 0.051N^{-1} + O(N^{-2}). \tag{4.7}$$

This last expression has been compared with the exact sum (4.3), and excellent agreement with the calculations of Montroll is obtained. There is, however, reason to distrust the third decimal of 0.1170 in (4.7).

For other square lattices one can derive expressions similar to (4.3). For a 3-dimensional simple cubic lattice, one obtains

$$D_3(n) = \frac{3}{2} \sum_{j_1 j_2 j_3}{}' \left[\sin^2\left(\frac{\pi j_1}{2n+1}\right) + \sin^2\left(\frac{\pi j_2}{2n+1}\right) + \sin^2\left(\frac{\pi j_3}{2n+1}\right) \right]^{-1}. \tag{4.8}$$

In this case one can, in order to find the asymptotic behavior of $D_3(n)$, replace the summation by an integration, which is now convergent. The result is

$$D_3(n) \simeq WN, \tag{4.9}$$

where W is Watson's integral

$$W = \frac{1}{(2\pi)^3} \int_{-\pi}^{\pi} \int\int \frac{dx\, dy\, dz}{1 - \tfrac{1}{3}(\cos x + \cos y + \cos z)} = 1.516386 \cdots. \tag{4.10}$$

Montroll[7] obtained formula (4.9) in a different way.

For the periodic square lattice, one can calculate the escape probability if at each step the exciton has the probability α of leaving the lattice. As in (3.4), this

escape probability for an exciton which starts at (x, y) satisfies the equations

$$E'_\alpha(x, y) = \alpha + \tfrac{1}{4}(1 - \alpha) \sum_{x', y'}^{(x,y)} E'_\alpha(x', y'), \quad (x, y) \neq (0, 0),$$

$$E'_\alpha(0, 0) = 0 \tag{4.11}$$

or, if $C_\alpha(x, y) = 1 - E'_\alpha(x, y)$ is the capture probability,

$$C_\alpha(x, y) = \tfrac{1}{4}(1 - \alpha) \sum_{x', y'}^{(x,y)} C_\alpha(x', y'), \quad (x, y) \neq (0, 0),$$

$$C_\alpha(0, 0) = 1. \tag{4.12}$$

The solution to these equations is written in the form

$$C_\alpha(x, y) = \sum_{j_1 j_2} A_\alpha(j_1, j_2) \cos\left(\frac{2\pi j_1 x}{2n + 1}\right) \cos\left(\frac{2\pi j_2 y}{2n + 1}\right), \quad \text{for all } (x, y), \tag{4.13}$$

and a closed expression for $A_\alpha(j_1, j_2)$ can be derived. One obtains

$$A_\alpha(j_1, j_2) = B_\alpha(j_1, j_2)/S_\alpha \tag{4.14}$$

with

$$S_\alpha = \sum_{j_1 j_2} B_\alpha(j_1, j_2) \tag{4.15}$$

and

$$B_\alpha(j_1, j_2) = \frac{1}{N}\left[1 - \frac{(1 - \alpha)}{2} \times \left(\cos\frac{2\pi j_1}{2n + 1} + \cos\frac{2\pi j_2}{2n + 1}\right)\right]^{-1}. \tag{4.16}$$

The probability of capture, averaged over all lattice points, is

$$C_\alpha \equiv N^{-1} \sum_{x, y} C_\alpha(x, y) = A_\alpha(0, 0) = (\alpha N S_\alpha)^{-1}. \tag{4.17}$$

For $\alpha = 0.002$ and $n = 9$ the values of $E'_\alpha(x, y)$, as obtained with (4.13)–(4.16), are given in Table V and should be compared with Table IV. The average escape probability $E'_\alpha(n) = 1 - C_\alpha(n)$ is plotted in Fig. 4 (upper curves) and must be compared with the corresponding lower curves for $E_\alpha(n)$. In both cases we took $n = 2$, 5, and 9.

TABLE V. The average probability of escape $E'_\alpha(x, y)$ for a periodic square lattice with $n = 9$ and $\alpha = 0.002$.

y										
9	0.624	0.625	0.627	0.631	0.636	0.640	0.645	0.649	0.651	0.653
8	0.619	0.620	0.623	0.627	0.632	0.638	0.643	0.647	0.650	
7	0.609	0.611	0.614	0.620	0.626	0.633	0.639	0.643		
6	0.594	0.596	0.601	0.608	0.617	0.625	0.633			
5	0.571	0.574	0.582	0.593	0.605	0.616				
4	0.539	0.544	0.557	0.573	0.590					
3	0.492	0.503	0.525	0.550						
2	0.420	0.446	0.486							
1	0.291	0.370								
0	0									
	0	1	2	3	4	5	6	7	8	9 x

The difference between the periodic and nonperiodic case is seen to be negligibly small. For all practical purposes, one can therefore calculate the average escape probability with the formula (4.17).

5. DISTRIBUTION OF TRAPPING CENTERS

Consider again the random walk on a square lattice with one trap. This trap, however, has not a definite position, but has the probability $P(n, m)$ of being at the lattice point (n, m). Let $D_{n,m}(x, y)$ be the average path length of an exciton, which starts at (x, y), with the trap fixed at the point (n, m). This average path length satisfies the equations [compare with (2.6)]

$$D_{n,m}(x, y) = 1 + \frac{1}{c(x, y)} \sum_{x', y'}^{(x,y)} D_{n,m}(x', y'), \quad (x, y) \neq (n, m),$$

$$D_{n,m}(n, m) = 0. \tag{5.1}$$

The quantity of interest is the path length averaged over the starting point and over the position of the trap, i.e.,

$$\bar{D} = \frac{1}{(2n + 1)^2} \sum_{x, y} \tilde{D}(x, y), \tag{5.2}$$

with

$$\tilde{D}(x, y) = \sum_{n, m} P(n, m) D_{n,m}(x, y). \tag{5.3}$$

On substitution one derives from (5.1) the following equations for $\tilde{D}(x, y)$, valid for all points (x, y):

$$\tilde{D}(x, y) = 1 - P(x, y)K(x, y) + \frac{1}{c(x, y)} \sum_{x', y'}^{(x,y)} \tilde{D}(x', y'), \tag{5.4}$$

with

$$K(x, y) = 1 + \frac{1}{c(x, y)} \sum_{x', y'}^{(x,y)} D_{x,y}(x', y'). \tag{5.5}$$

The last equation shows that $K(x, y)$ may be interpreted as the average path-length traveled by an exciton, when it returns for the first time to the starting point (x, y). Using this interpretation, we will separably calculate the function $K(x, y)$ later in this section.

Turning to Eq. (5.4), we should observe that $\tilde{D}(x, y)$ cannot be solved uniquely from it, since an arbitrary constant can be added to the solution. An auxiliary function $F(x, y)$ is therefore defined by the requirements that $F(x, y)$ satisfy (5.4) for all (x, y) and $F(0, 0) = 0$. Then

$$\tilde{D}(x, y) = F(x, y) + \tilde{D}(0, 0). \tag{5.6}$$

$\tilde{D}(x, y)$ can be calculated in the following steps:

(i) Calculate $\tilde{D}(0, 0)$ from $D_{n,m}(0, 0)$ with Eq. (5.3).

(ii) Find $K(x, y)$ for all (x, y).

(iii) Solve

$$F(x, y) = 1 - P(x, y)K(x, y) + \frac{1}{c(x, y)} \sum_{x', y'}^{(x,y)} F(x', y')$$

$$\text{for all } (x, y) \text{ together with } F(0, 0) = 0. \quad (5.7)$$

As an auxiliary theorem to the first point, we now show that

$$D_{n,m}(0, 0) = \tfrac{1}{4}c(n, m)D_{0,0}(n, m). \quad (5.8)$$

Both $D_{n,m}(0, 0)$ and $D_{0,0}(n, m)$ must be calculated as an average over all possible paths P connecting (n, m) and $(0, 0)$.

Let $W(P)$ be the probability for such a path, when it is traversed from (n, m) to $(0, 0)$, and $W'(P)$, when it is traversed in the opposite direction. Each of these probabilities can be written as the product of a number of factors $\frac{1}{4}$, $\frac{1}{3}$, and $\frac{1}{2}$:

$$W(P) = (\tfrac{1}{2})^a(\tfrac{1}{3})^b(\tfrac{1}{4})^c$$

and

$$W'(P) = (\tfrac{1}{2})^{a'}(\tfrac{1}{3})^{b'}(\tfrac{1}{4})^{c'},$$

where a is the number of steps from a corner point to a neighbor, b is the number of steps from a side point to a neighbor, c is the number of steps from an interior point to a neighbor, and a', b', and c' are defined in the same way, but for the path $(0, 0) \to (n, m)$.

It is easy to convince oneself that: If (n, m) is an interior point, then $a' = a$, $b' = b$, $c' = c$; if (n, m) is a side point, then $a' = a$, $b' = b - 1$, $c' = c + 1$; if (n, m) is a corner point, then $a' = a - 1$, $b' = b$, $c' = c + 1$. From this it follows that for any path P, connecting $(0, 0)$ with an arbitrary point (n, m), one has

$$W'(P) = \tfrac{1}{4}c(n, m)W(P).$$

This result immediately implies (5.8).

Since $D_{0,0}(n, m)$ is the average path length to reach the point $(0, 0)$ from (n, m), it equals the quantity $D^{(2)}(x, y)$ from Sec. 2. This quantity can be evaluated numerically from Eq. (2.6), as has been done for a few cases (Table II). This calculation is independent of the functions $P(n, m)$.

Since with Eqs. (5.3) and (5.8)

$$\tilde{D}(0, 0) = \tfrac{1}{4} \sum_{n,m} c(n, m)P(n, m)D_{0,0}(n, m), \quad (5.9)$$

the average value $\tilde{D}(0, 0)$ can be found numerically for any given distribution P.

As the next step in solving Eq. (5.7), the average path length $K(x, y)$ upon the first return to the point (x, y) must be calculated. Consider the random walk of a particle on a lattice, and let it take a large number of steps, say M. Let $n_M(x, y)$ be the number of times the particle is found at the point (x, y). Then one has

$$K(x, y) = \lim_{M \to \infty} \frac{M}{n_M(x, y)}. \quad (5.10)$$

For a very long walk the probability $p(x, y)$ of finding the particle at the point (x, y) becomes independent of the starting point, and therefore

$$p(x, y) = \lim_{M \to \infty} \frac{n_M(x, y)}{M}, \quad (5.11)$$

so that

$$K(x, y) = \frac{1}{p(x, y)}.$$

For $M \to \infty$, the probability of finding the particle at the point (x, y) is stationary, and $p(x, y)$ satisfies the equation

$$p(x, y) = \sum_{x', y'}^{(x,y)} \frac{p(x', y')}{c(x', y')}. \quad (5.12)$$

This equation expresses how $p(x, y)$ can be calculated in terms of the probabilities of finding the particle at the neighboring points. The solution of (5.12) is

$$p(x, y) = p_0 c(x, y),$$

where the constant p_0 is determined by the requirement

$$\sum_{x,y} p(x, y) = 1.$$

This gives

$$p_0^{-1} = \sum_{x,y} c(x, y) = 8(2n^2 + n).$$

Thus

$$K(x, y) = \frac{1}{c(x, y)} \sum_{x', y'} c(x', y')$$

$$= \frac{8}{c(x, y)} (2n^2 + n)$$

$$= \frac{4}{c(x, y)} (N - \sqrt{N}), \quad (5.13)$$

in which N is the total number of lattice points in the lattice: $N = (2n + 1)^2$. It further appears that, because it is obvious that $K(0, 0) = D_{0,0}(1, 0) + 1$, Eq. (5.13) also is a proof of Eq. (2.7).

As the last step in determining $\tilde{D}(x, y)$, we can now solve $F(x, y)$ from Eq. (5.7) in a numerical way with the aid of Eq. (5.13). Adding the numerical result of

TABLE VI. The average path length $\tilde{D}(x, y)$ for a square lattice with $n = 9$ and a trap distribution with $r = 6$ [cf. Eq. (5.14)].

y										
9	705	706	708	712	717	722	727	731	734	734
8	701	702	705	709	715	720	725	730	732	
7	694	695	698	703	709	715	721	726		
6	684	685	689	695	701	709	716			
5	672	674	678	684	692	701				
4	660	661	666	674	683					
3	648	650	655	663						
2	638	640	646							
1	632	634								
0	630									
	0	1	2	3	4	5	6	7	8	9 x

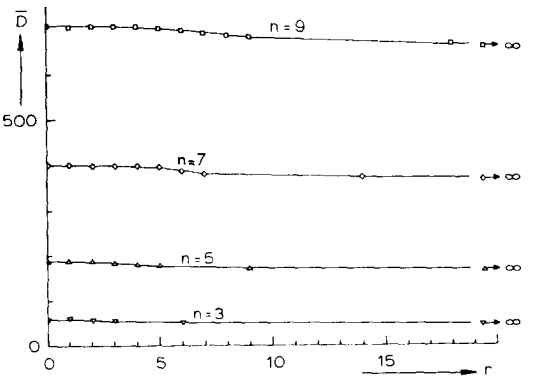

FIG. 5. The average path length as a function of the range r of the trap distribution [cf. Eq. (5.14)].

Eq. (5.9) to $F(x, y)$, we obtain the unique solution for $\tilde{D}(x, y)$, which gives the average path length $\bar{D}$ according to Eq. (5.2).

This sequence of steps (i)–(iii) has been performed for n from 2 to 9 with the distribution

$$P(n, m) = \exp\left(-\frac{n^2 + m^2}{r^2}\right)\left[\sum_{n',m'} \exp\left(-\frac{n'^2 + m'^2}{r^2}\right)\right]^{-1}. \tag{5.14}$$

This distribution of traps is symmetric around the center of the lattice and has a range r. A uniform distribution of traps is obtained for $r \gg n$. It was somewhat of a surprise to find that for such a uniform trap distribution the path length averaged over all trap positions, i.e., $\tilde{D}(x, y)$, was not quite constant. For $n = 9$ and $r = \infty$ it varied between 656.06 for a corner point and 664.59 for the center.

In Table VI we give $\tilde{D}(x, y)$ for $n = 9$ and $r = 6$. The effect of distributing the trapping centers can be judged by comparison with Table II. Another way is to plot $\bar{D}$, the average value of $\tilde{D}(x, y)$ [Eq. (5.2)], as a function of the range parameter r of the distribution function $P(n, m)$. For a number of values of n the result can be seen in Fig. 5. It is seen that, although $\tilde{D}(x, y)$ may vary considerably with r, $\bar{D}$ changes very little with r going from zero to infinity.

We finally note that the calculation of $K(x, y)$ in step (ii) gives rise to a general theorem. Consider a connected, but otherwise arbitrary finite network of points. Let the point λ be connected with $c(\lambda)$ other points of the network. Then the average number of steps a random walker takes, before he returns to his starting point λ, is, according to (5.13), equal to

$$K(\lambda) = \frac{1}{c(\lambda)} \sum_{\mu} c(\mu). \tag{5.15}$$

Although this result probably can be found in the literature, it was perhaps worthwhile to give this simple proof.

[1] J. Avery, Z. Bay, and A. Szent-Györgyi, Proc. Natl. Acad. Sci. (U.S.) **47**, 1742 (1961).
[2] J. J. ten Bosch and Th. W. Ruijgrok, J. Theoret. Biol. **4**, 225 (1963).
[3] R. M. Pearlstein, thesis, University of Maryland, 1966; Brookhaven Natl. Lab. Symp. **19**, 19 (1967).
[4] G. W. Robinson, Brookhaven Natl. Lab. Symp. **19**, 16 (1967).
[5] R. S. Knox, J. Theoret. Biol. **21**, 244 (1968).
[6] E. W. Montroll, Proc. Symp. Appl. Math. **16**, 193 (1964). See especially p. 206.
[7] E. W. Montroll, J. Math. Phys. **10**, 753 (1969).
[8] L. N. M. Duysens, Progr. Biophys. Mol. Biol. **14**, 1 (1964).

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 12, NUMBER 3 MARCH 1971

Eigenvalue Moments for Ensembles of Random Matrices*

J. F. McDONALD
Department of Mathematics, University of Windsor, Windsor, Ontario, Canada

(Received 14 September 1970)

A straightforward method is given for calculating the moments of the joint eigenvalue distribution corresponding to an ensemble of random Hermitian matrices. The method enables one to calculate the averages using the matrix element distribution. Thus an explicit expression for the joint eigenvalue distribution is not required.

1. INTRODUCTION

In recent years there has been considerable success in predicting the statistical properties of the high-lying energy levels of complex systems using ensembles of random matrices.[1] Most of the work to date has been based on ensembles which are representationally invariant (i.e., the joint distribution for the matrix elements, which mathematically specifies the ensemble, is a function only of traces of various powers of the Hamiltonian matrix). For ensembles of this type the joint eigenvalue distribution and the distribution of widths can be obtained explicitly.[2–6]

Recently there has been interest in the problem of how a small time-reversal odd term in the Hamiltonian of a complex system would influence the statistical properties of the energy levels.[7–14] Such investigations have led to interest in ensembles which are not representationally invariant.

For such ensembles (with the dimensionality of the matrices arbitrary) it does not seem feasible to obtain exact expressions for the various distributions involving the eigenvalues and widths. Thus most of what has been done involves approximations and numerical calculations.[7–14]

However, we shall show that, even if the joint eigenvalue distribution is unknown, its moments can be calculated exactly using the matrix element distribution.

2. MATHEMATICAL DEFINITION OF ENSEMBLES AND ENSEMBLE AVERAGES

Consider an ensemble of $N \times N$ Hermitian matrices. We shall denote a typical member of the ensemble as H. The real and imaginary parts of H will be denoted by R and S, respectively. Since H is Hermitian,

$$R_{ij} = R_{ji}, \tag{1}$$

$$S_{ij} = -S_{ji}, \quad i \neq j, \tag{2}$$

and

$$S_{ii} = 0. \tag{3}$$

The ensemble can be specified mathematically by the joint distribution for the N^2 independent variables, R_{ij}, $i \geq j$, and S_{ij}, $i > j$. We shall denote this joint matrix element distribution as $p(H)$. The volume element in this N^2-dimensional space will be denoted by dH, where

$$dH = \prod_{i \geq j} dR_{ij} \prod_{k > l} dS_{kl}. \tag{4}$$

All of the variables can be assumed to vary independently from $-\infty$ to $+\infty$.

The properties of the physical system being considered can be used to impose certain restrictions on the matrix element distribution.[5,15] However, at this point the only restriction we impose on $p(H)$ is that it is normalized to unity.

To obtain the joint eigenvalue distribution corresponding to $p(H)$, one first makes a change of variables from the N^2 variables (R_{ij}, S_{ij}) to the N eigenvalues E_i, $i = 1, \cdots, N$, and $N(N-1)$ other variables, say ϕ_α, $\alpha = 1, 2, \cdots, N(N-1)$.[4,5] Thus, we have

$$p(H)\, dH = p(E, \phi) J(R, S: E, \phi)\, dE\, d\phi, \tag{5}$$

where $p(E, \phi)$ is $p(H)$ expressed in terms of the new variables, $J(R, S: E, \phi)$ is the Jacobian of the transformation,

$$dE \equiv \prod dE_i, \tag{6}$$

and

$$d\phi \equiv \prod d\phi_i. \tag{7}$$

The joint eigenvalue distribution $p(E)$ is then obtained by integration over the ϕ_i, i.e.,

$$p(E) = \int d\phi\, p(E, \phi) J(R, S: E, \phi). \tag{8}$$

If $Q(E)$ is any function of the eigenvalues, its ensemble average $\langle Q \rangle$ is defined as

$$\langle Q \rangle = \int p(E) Q(E)\, dE. \tag{9}$$

It is clear that

$$\int p(E)Q(E)\,dE = \int dE \int d\phi J p(E,\phi)Q(E) = \int p(H)Q(H)\,dH, \tag{10}$$

where $Q(H)$ is $Q(E)$ expressed in terms of the variables R_{ij} and S_{ij}. Thus, the ensemble average of a function of the eigenvalues can be obtained either by using the joint eigenvalue distribution directly or by expressing the quantity of interest in terms of the matrix elements and averaging using the matrix element distribution.

In what follows it will be assumed that $p(E)$ is the symmetrized joint eigenvalue distribution. That is, $p(E)$ is invariant with respect to any permutation of the indices on the E_i. Thus, $\langle Q(E)\rangle$ is invariant to any permutation of the indices on the E_i in the expression for $Q(E)$. In particular,

$$\langle E_{i_1}^{n_1} E_{i_2}^{n_2} \cdots E_{i_m}^{n_m}\rangle = \langle E_1^{n_1} E_2^{n_2} \cdots E_m^{n_m}\rangle, \tag{11}$$

provided that the i_α, $\alpha = 1, \cdots, m$, are distinct.

3. EIGENVALUE MOMENTS IN TERMS OF AVERAGES OF TRACES OF THE HAMILTONIAN MATRIX

As shown above, ensemble averages of functions of the eigenvalues can be evaluated using the joint matrix element distribution if the function can be written in terms of the variables R_{ij} and S_{ij}. In general, this is not possible since the explicit functional dependence of the eigenvalues on the matrix elements for an $N \times N$ matrix is not known. However, in some cases the explicit functional relationship is not required. In particular, for quantities of the form

$$Q(E) = \prod_{\alpha=1}^{m} E_\alpha^{n_\alpha}, \tag{12}$$

where the n_α are nonnegative integers, the relationship

$$\operatorname{Tr} H^n = \sum_i E_i^n, \tag{13}$$

where n is a nonnegative integer and

$$\operatorname{Tr} H^0 \equiv N, \tag{14}$$

is sufficient.

To evaluate averages of this form, we consider the ensemble average of the quantity

$$T(H) \equiv \prod_{\alpha=1}^{m} \operatorname{Tr} H^{n_\alpha}. \tag{15}$$

From Eq. (13) it trivially follows that

$$\langle T(H)\rangle = \sum_{i_1,\cdots,i_m} \langle E_{i_1}^{n_1} \cdots E_{i_m}^{n_m}\rangle. \tag{16}$$

The m summations on the right-hand side can be rewritten as

$$\sum_{i_1,\cdots,i_m} = \sum_{i_1=i_2=\cdots=i_m} + \sum'_{i_1,i_2=\cdots=i_m} + \sum'_{i_2,i_1=i_3=\cdots=i_m} + \cdots + \sum'_{i_m,i_1=\cdots=i_{m-1}} + \sum'_{i_1=i_2,i_3=i_4=\cdots=i_m} + \cdots + \sum'_{i_{m-1}=i_m,i_1=\cdots=i_{m-2}} + \cdots + \sum'_{i_1,i_2,\cdots,i_m}, \tag{17}$$

where the prime on a summation sign means that only terms where the remaining indices are all unequal are to be included. Further, from Eq. (11) it is seen that each term in any of the new sums is identical. Thus, the summation signs just introduce a multiplicative factor η_α, where $\alpha + 1$ is the number of remaining summation indices and

$$\eta_\alpha \equiv N(N-1)\cdots(N-\alpha). \tag{18}$$

Thus Eq. (16) can be rewritten as

$$\langle T(H)\rangle = N\langle E^n\rangle + \eta_1 \sum_{\alpha=1}^{m}{}^{D} \langle E_1^{n_\alpha} E_2^{n-n_\alpha}\rangle + \eta_1 \sum_{\alpha>\beta}^{D} \langle E_1^{n_\alpha+n_\beta} E_2^{n-n_\alpha-n_\beta}\rangle + \cdots + \eta_{m-1}\langle E_1^{n_1} E_2^{n_2} \cdots E_m^{n_m}\rangle, \tag{19}$$

where

$$n = \sum_{\alpha=1}^{m} n_\alpha \tag{20}$$

and where $\sum^D$ means that the sum is to include only "distinct" terms.[16]

From Eq. (19) it can be seen that if $\langle E_1^{s_1} E_2^{s_2} \cdots E_t^{s_t}\rangle$, $t = 1, 2, \cdots, m-1$, are known in terms of averages of traces of H, for arbitrary nonnegative integers S_α, then $\langle E_1^{n_1} E_2^{n_2} \cdots E_m^{n_m}\rangle$ can easily be found for arbitrary n_α in terms of averages of traces of H.[17] Thus, it is clear that $\langle E_1^{n_1} E_2^{n_2} \cdots E_m^{n_m}\rangle$ can be found by an iterative procedure.

We begin by considering $m = 1$. Now

$$\langle \operatorname{Tr} H^n\rangle = \sum_i \langle E_i^n\rangle = N\langle E_1^n\rangle, \tag{21}$$

or

$$\langle E_1^n\rangle = \frac{1}{N}\langle \operatorname{Tr} H^n\rangle. \tag{22}$$

Proceeding to $m = 2$, we have

$$\langle \operatorname{Tr} H^{n_1} \operatorname{Tr} H^{n_2}\rangle = \sum_{i,j} \langle E_i^{n_1} E_j^{n_2}\rangle = N\langle E_1^n\rangle + \eta_1\langle E_1^{n_1} E_2^{n_2}\rangle. \tag{23}$$

From Eqs. (22) and (23) it then follows that

$$\langle E_1^{n_1} E_2^{n_2}\rangle = \eta_1^{-1}\langle \operatorname{Tr} H^{n_1} \operatorname{Tr} H^{n_2}\rangle', \tag{24}$$

where

$$\langle \mathrm{Tr}\, H^{n_1}\, \mathrm{Tr}\, H^{n_2}\rangle' \equiv \langle \mathrm{Tr}\, H^{n_1}\, \mathrm{Tr}\, H^{n_2}\rangle - \langle \mathrm{Tr}\, H^{n}\rangle. \quad (25)$$

For $m = 3$

$$\begin{aligned}\mathrm{Tr}\, H^{n_1}\, \mathrm{Tr}\, H^{n_2}\, \mathrm{Tr}\, H^{n_3}\rangle &= \sum_{i,j,k} \langle E_i^{n_1} E_j^{n_2} E_k^{n_3}\rangle \\ &= N\langle E_1^{n}\rangle + \eta_1 \sum_{\alpha=1}^{3} \langle E_1^{n_\alpha} E_2^{n-n_\alpha}\rangle \\ &\quad + \eta_2 \langle E_1^{n_1} E_2^{n_2} E_3^{n_3}\rangle. \quad (26)\end{aligned}$$

Combining equations (22), (24), and (26), we obtain

$$\langle E_1^{n_1} E_2^{n_2} E_3^{n_3}\rangle = \frac{1}{\eta_2}\left(\langle \mathrm{Tr}\, H^{n_1}\, \mathrm{Tr}\, H^{n_2}\, \mathrm{Tr}\, H^{n_3}\rangle' - \sum_{\alpha=1}^{3} \langle \mathrm{Tr}\, H^{n_\alpha}\, \mathrm{Tr}\, H^{n-n_\alpha}\rangle'\right). \quad (27)$$

Obviously this procedure can be continued indefinitely. For example, the reader can verify that

$$\begin{aligned}\langle E_1^{n_1} E_2^{n_2} E_3^{n_3} E_4^{n_4}\rangle = \eta_3^{-1}\Big(&\langle \mathrm{Tr}\, H^{n_1}\, \mathrm{Tr}\, H^{n_2}\, \mathrm{Tr}\, H^{n_3}\, \mathrm{Tr}\, H^{n_4}\rangle' \\ &- \sum_{\alpha>\beta=1}^{4} \langle \mathrm{Tr}\, H^{n_\alpha}\, \mathrm{Tr}\, H^{n_\beta}\, \mathrm{Tr}\, H^{n-n_\alpha-n_\beta}\rangle' \\ &+ 2\sum_{\alpha=1}^{4} \langle \mathrm{Tr}\, H^{n_\alpha}\, \mathrm{Tr}\, H^{n-n_\alpha}\rangle' \\ &+ \sum_{\alpha=2}^{4} \langle \mathrm{Tr}\, H^{n_1+n_\alpha}\, \mathrm{Tr}\, H^{n-n_1-n_\alpha}\rangle'\Big). \quad (28)\end{aligned}$$

In all of the above $\langle T(H)\rangle' \equiv \langle T(H)\rangle - \langle \mathrm{Tr}\, H^{n}\rangle$.

At this point, one might try to generalize to the case where m is arbitrary. First of all it follows from mathematical induction that

$$\langle E_1^{n_1} E_2^{n_2} \cdots E_m^{n_m}\rangle = \frac{1}{\eta_{m-1}} \sum_{\alpha=2}^{m} \sum_{s_1, s_2, \cdots, s_\alpha} C_{s_1 s_2 \cdots s_\alpha} \langle \mathrm{Tr}\, H^{s_1} \cdots \mathrm{Tr}\, H^{s_\alpha}\rangle', \quad m \geq 2, \quad (29)$$

where the sum includes only those $(s_1, s_2, \cdots, s_\alpha)$ such that each s_i is a sum of the n_i and each n_i occurs once and only once. The sum is to include only sets of s_i for which $\langle \mathrm{Tr}\, H^{s_1} \cdots \mathrm{Tr}\, H^{s_\alpha}\rangle'$ are distinct for arbitrary n_i. For example, if $s_1 = n_1$, $s_2 = n - n_1$ is included, then $s_1 = n - n_1$, $s_2 = n_1$ need not be included. Of course, for a particular set of n_i some of the terms in the sum may be identical. For example, $(n_1, n_2 + n_3, n - n_1 - n_2 - n_3)$ and $(n_2, n_1 + n_3, n - n_1 - n_3)$ are distinct for arbitrary n_i and are thus both included in the sum. However, if $n_1 = n_2$, the two averages will be identical.

Some properties of the coefficients $C_{s_1 \cdots s_\alpha}$ can easily be deduced. First of all, the coefficients are independent of n. Second the coefficients depend only on m and not on the particular values of the n_i. Both of these can be established using mathematical induction.[18,19] Finally, since the result is invariant with respect to a permutation of n_i, terms of the same type will have the same coefficient.[20] Thus, for example,

$$C_{n_1, n_2+n_3, n_4+n_5+n_6} = C_{n_2, n_4+n_1, n_3+n_5+n_6}. \quad (30)$$

The explicit calculation of the coefficients for the general case is not trivial. Basically the problem of determining them is a counting problem. One must trace through the origin of such terms, taking into account the appropriate sign and numerical coefficients. A general procedure for this can be formulated, but it is cumbersome. Thus, we shall not attempt to state it here.

Of course, for any particular case, i.e., fixed m, the required coefficients can be obtained as above for $m = 1, 2, 3$, and 4. This, of course, becomes more and more laborious as m increases. There is another approach which seems preferable.[21]

One simply writes out (29) for the case of interest, where the C's are to be treated as unknowns. Since the C's are independent of the n, one then chooses special cases, i.e., chooses particular values for the n. If there are m independent C's, one must be careful to choose m special cases which give independent equations. The problem thus reduces to the solution of m simultaneous linear equations.

Two particular cases which are not included in the above results are given here for reference. They are

$$\begin{aligned}\langle E_1 E_2 E_3 E_4 E_5^2\rangle &= \eta_4^{-1}[\langle (\mathrm{Tr}\, H)^4\, \mathrm{Tr}\, H^2\rangle' - 24\langle \mathrm{Tr}\, H\, \mathrm{Tr}\, H^5\rangle' \\ &\quad - 18\langle \mathrm{Tr}\, H^2\, \mathrm{Tr}\, H^4\rangle' - 8\langle (\mathrm{Tr}\, H^3)^2\rangle' \\ &\quad + 12\langle (\mathrm{Tr}\, H)^2\, \mathrm{Tr}\, H^4\rangle' + 20\langle \mathrm{Tr}\, H\, \mathrm{Tr}\, H^2\, \mathrm{Tr}\, H^3\rangle' \\ &\quad + 3\langle (\mathrm{Tr}\, H^2)^3\rangle' - 6\langle (\mathrm{Tr}\, H)^2 (\mathrm{Tr}\, H^2)^2\rangle' \\ &\quad - 4\langle (\mathrm{Tr}\, H)^3 (\mathrm{Tr}\, H^3)\rangle'], \quad (31)\end{aligned}$$

and

$$\begin{aligned}\langle E_1 E_2 E_3 E_4 E_5 E_6\rangle &= \eta_5^{-1}[\langle (\mathrm{Tr}\, H)^6\rangle' + 144\langle \mathrm{Tr}\, H\, \mathrm{Tr}\, H^5\rangle' \\ &\quad + 90\langle \mathrm{Tr}\, H^2\, \mathrm{Tr}\, H^4\rangle' + 40\langle (\mathrm{Tr}\, H^3)^2\rangle' \\ &\quad - 90\langle (\mathrm{Tr}\, H)^2\, \mathrm{Tr}\, H^4\rangle' - 120\langle \mathrm{Tr}\, H\, \mathrm{Tr}\, H^2\, \mathrm{Tr}\, H^3\rangle' \\ &\quad - 15\langle (\mathrm{Tr}\, H^2)^3\rangle' + 40\langle (\mathrm{Tr}\, H)^3\, \mathrm{Tr}\, H^3\rangle' \\ &\quad + 45\langle (\mathrm{Tr}\, H)^2 (\mathrm{Tr}\, H^2)^2\rangle' - 15\langle (\mathrm{Tr}\, H)^4\, \mathrm{Tr}\, H^2\rangle'], \quad (32)\end{aligned}$$

These two results will be used in the following sections.

4. EIGENVALUE MOMENTS IN TERMS OF AVERAGES OF THE MATRIX ELEMENTS

Since Tr H^n can be written down explicitly in terms of the R_{ij} and S_{ij}, for any particular n, where n is a nonnegative integer, it is clear that, in principle, $\langle E_1^{n_1} \cdots E_m^{n_m} \rangle$ can be evaluated for a particular set of n_i if $p(H)$ is given. To this point, $p(H)$ was completely arbitrary. For simplicity, we shall restrict our discussion from this point to distributions of the form

$$p(H) = \prod_i f(R_{ii}) \prod_{j>k} g(R_{jk}) h(S_{jk}). \tag{33}$$

Note that all of the variables are statistically independent. Further, all R_{ii}, $i = 1, \cdots, N$, have the same distribution. Similarly, for the R_{jk} and S_{jk}. Finally, we assume that

$$\langle R_{ii}^{2n+1} \rangle = \langle R_{jk}^{2n+1} \rangle = \langle S_{jk}^{2n+1} \rangle = 0, \quad n = 0, 1, 2, \cdots. \tag{34}$$

It follows immediately from this that

$$\left\langle \prod_\alpha H_{j_\alpha k_\alpha}^{n_\alpha} \right\rangle = 0 \tag{34'}$$

if $\sum_\alpha n_\alpha$ is odd. It should be noted that most of the ensembles considered to date are members of this restricted class, defined by (33) and (34).

In addition, it is easily seen that for ensembles contained in the above class

$$\langle R_{i_1 i_1}^{n_1} \cdots R_{i_r i_r}^{n_r} R_{j_1 k_1}^{p_1} \cdots R_{j_s k_s}^{p_s} S_{l_1 m_1}^{q_1} \cdots S_{l_t m_t}^{q_t} \rangle = \prod_{i=1}^{r} \langle x^{n_i} \rangle \prod_{j=1}^{s} \langle \{\mathrm{Re}\ y\}^{p_j} \rangle \prod_{k=1}^{t} \langle \{\mathrm{Im}\ y\}^{q_k} \rangle, \tag{35}$$

and

$$\langle H_{i_1 i_1}^{n_1} \cdots H_{i_r i_r}^{n_r} H_{j_1 k_1}^{p_1} |H_{j_1 k_1}|^{q_1} \cdots H_{j_s k_s}^{p_s} |H_{j_s k_s}|^{q_s} \rangle = \prod_{i=1}^{r} \langle x^{n_i} \rangle \prod_{j=1}^{s} \langle y^{p_i} |y|^{q_i} \rangle, \tag{35'}$$

where x and y denote a typical diagonal element, say H_{11}, and a typical off-diagonal element, say H_{21}, respectively. It has, of course, been assumed that the i_α, $\alpha = 1, \cdots, r$, are distinct, the pairs (j_α, k_α), $j_\alpha > k_\alpha$, $\alpha = 1, \cdots, s$, are distinct, and the pairs (l_α, m_α), $l_\alpha > m_\alpha$, $\alpha = 1, \cdots, t$, are distinct.

Consider first $\langle E_1^{n_1} E_2^{n_2} \cdots E_m^{n_m} \rangle$, where $n = \sum n_i$ is odd. Since $\sum_{i=1}^{\alpha} s_i = n$, $\alpha = 2, \cdots, m$, for those terms included in the sum in Eq. (29), and since

$$\mathrm{Tr}\, H^{s_i} = \sum_{j_1, j_2, \cdots, j_{s_i}} H_{j_1 j_2} H_{j_2 j_3} \cdots H_{j_{s_i - 1} j_{s_i}} H_{j_{s_i} j_1}, \tag{36}$$

it is clear that each term in the sum over the s_i is a sum of terms each of which is a product of the n matrix elements. Thus, from Eq. (34') it follows that every term vanishes. Thus,

$$\langle E_1^{n_1} \cdots E_m^{n_m} \rangle = 0 \tag{37}$$

if n is odd.

Consider next the case when n is even. The evaluation of $\langle E_1^{n_1} \cdots E_m^{n_m} \rangle$ is very straightforward. One simply inserts the expressions for Tr H^{s_α} [Eq. (36)] into Eq. (29), rewrites it in terms of the R_{ii}, R_{ij}, and S_{ij} and evaluates the average term by term. Many terms vanish and many are equal. These results follows from (34) and (35), respectively.

However, it is clear that the complexity of this calculation increases as m increases because the number of averages required likewise increases. In addition, as the n_i increase, the traces involve higher powers of H. As the power increases, the corresponding expression for the trace in terms of the matrix elements becomes more complicated. Thus, the general case seems difficult at best. Thus, for definiteness we shall restrict our discussion to the cases $n = 2$, 4, and 6.

In Tables I–III the averages of the traces required to evaluate all of the eigenvalue moments with $n = 2$, 4, and 6 are given. The left-hand column indicates the quantity being averaged, the first row indicates the matrix element averages required, and the other entries denote the corresponding coefficients. Thus, for example, from Table II we have

$$\langle \mathrm{Tr}\, H^4 \rangle = N\langle x^4 \rangle + 4\eta_1 \langle x^2 \rangle \langle |y|^2 \rangle + \eta_1 \langle |y|^4 \rangle + 2\eta_2 \{\langle |y|^2 \rangle\}^2; \tag{38}$$

the other averages can be read off the tables in a similar manner.

The expressions for all $\langle E_1^{n_1} \cdots E_m^{n_m} \rangle$ with $n = 2$, 4, and 6 can now be obtained by inserting the results given in Tables I–III into Eqs. (22), (24), (27),

TABLE I. Averages of traces, required for eigenvalue moments with $n = 2$, expressed in terms of averages of matrix elements.

	$\langle x^2 \rangle$	$\langle \lvert y \rvert^2 \rangle$
$\langle \mathrm{Tr}\, H^2 \rangle$	N	η_1
$\langle (\mathrm{Tr}\, H)^2 \rangle$	N	0

TABLE II. Averages of traces, required for eigenvalue moments with $n = 4$, expressed in terms of averages of matrix elements.

	$\langle x^4 \rangle$	$\{\langle x^2 \rangle\}^2$	$\{\langle x^2 \rangle \langle \lvert y \rvert^2 \rangle\}$	$\langle \lvert y \rvert^4 \rangle$	$\{\langle \lvert y \rvert^2 \rangle\}^2$
$\langle \mathrm{Tr}\, H^4 \rangle$	N	0	$4\eta_1$	η_1	$2\eta_2$
$\langle \mathrm{Tr}\, H^3\, \mathrm{Tr}\, H \rangle$	N	0	$3\eta_1$	0	0
$\langle (\mathrm{Tr}\, H^2)^2 \rangle$	N	η_1	$2N\eta_1$	$2\eta_1$	$(N+1)\eta_2$
$\langle \mathrm{Tr}\, H^2 (\mathrm{Tr}\, H)^2 \rangle$	N	η_1	$N\eta_1$	0	0
$\langle \mathrm{Tr}\, H^4 \rangle$	N	$3\eta_1$	0	0	0

TABLE III. Averages of traces, required for eigenvalue moments with $n = 6$, expressed in terms of averages of matrix elements.

	$\langle x^6\rangle$	$\langle x^4\rangle\langle x^2\rangle$	$\{\langle x^2\rangle\}^3$	$\langle x^4\rangle\langle\|y\|^2\rangle$	$\{\langle x^2\rangle\}^2\langle\|y\|^2\rangle$	$\langle x^2\rangle\langle\|y\|^4\rangle$	$\langle x^2\rangle\{\langle\|y\|^2\rangle\}^2$	$\langle\|y\|^6\rangle$	$\langle\|y\|^4\rangle\langle\|y\|^2\rangle$	$\{\langle\|y\|^2\rangle\}^3$	$\{\langle y^2\rangle\}^3$
$\langle \mathrm{Tr}\, H^6\rangle$	N	0	0	$6\eta_1$	$3\eta_1$	$9\eta_1$	$15\eta_2$	η_1	$6\eta_2$	$(5N-12)\eta_2$	η_2
$\langle \mathrm{Tr}\, H^5\,\mathrm{Tr}\, H\rangle$	N	0	0	$5\eta_1$	$5\eta_1$	$5\eta_1$	$10\eta_2$	0	0	0	0
$\langle \mathrm{Tr}\, H^4\,\mathrm{Tr}\, H^2\rangle$	N	η_1	0	$(N+4)\eta_1$	$4(N-1)\eta_1$	$(N+8)\eta_1$	$2(3N+2)\eta_2$	$2\eta_1$	$(N+9)\eta_2$	$2(N^2-N-4)\eta_2$	0
$\langle \mathrm{Tr}\, H^4(\mathrm{Tr}\, H)^2\rangle$	N	η_1	0	$4\eta_1$	$4N\eta_1$	$N\eta_1$	$2N\eta_2$	0	0	0	0
$\langle(\mathrm{Tr}\, H^3)^2\rangle$	N	0	0	$6\eta_1$	0	$9\eta_1$	$9\eta_2$	0	0	$3\eta_2$	$3\eta_2$
$\langle \mathrm{Tr}\, H^3\,\mathrm{Tr}\, H^2\,\mathrm{Tr}\, H\rangle$	N	η_1	0	$(N+3)\eta_1$	$3(N-1)\eta_1$	$6\eta_1$	$3(N+1)\eta_2$	0	0	0	0
$\langle \mathrm{Tr}\, H^3(\mathrm{Tr}\, H)^3\rangle$	N	$3\eta_1$	0	$3\eta_1$	$9(N-1)\eta_1$	0	0	0	0	0	0
$\langle(\mathrm{Tr}\, H^2)^3\rangle$	N	$3\eta_1$	η_2	$3N\eta_1$	$3\eta_1^2$	$6N\eta_1$	$3N(N+1)\eta_2$	$4\eta_1$	$6(N+1)\eta_2$	$(N+1)\eta_2(\eta_1-4)$	0
$\langle(\mathrm{Tr}\, H^2)^2(\mathrm{Tr}\, H)^2\rangle$	N	$3\eta_1$	η_2	$2N\eta_1$	$2\eta_1^2$	$2N\eta_1$	$N(N+1)\eta_2$	0	0	0	0
$\langle \mathrm{Tr}\, H^2(\mathrm{Tr}\, H)^4\rangle$	N	$7\eta_1$	$3\eta_2$	$N\eta_1$	$3\eta_1^2$	0	0	0	0	0	0
$\langle(\mathrm{Tr}\, H)^6\rangle$	N	$15\eta_1$	$15\eta_2$	0	0	0	0	0	0	0	0

TABLE VI. Eigenvalue moments, with $n = 6$, expressed in terms of averages of matrix elements.

	$\langle x^6\rangle$	$\langle x^4\rangle\langle x^2\rangle$	$\{\langle x^2\rangle\}^3$	$\langle x^4\rangle\langle\|y\|^2\rangle$	$\{\langle x^2\rangle\}^2\langle\|y\|^2\rangle$	$\langle x^2\rangle\langle\|y\|^4\rangle$	$\langle x^2\rangle\{\langle\|y\|^2\rangle\}$	$\langle\|y\|^6\rangle$	$\langle\|y\|^4\rangle\langle\|y\|^2\rangle$	$\{\langle\|y\|^2\rangle\}^3$	$\{\langle y^2\rangle\}^3$
$\langle E_1^6\rangle$	1	0	0	$6N_1$	$3N_1$	$9N_1$	$15N_1N_2$	N_1	$6N_1N_2$	$(5N-12)N_1N_2$	N_1N_2
$\langle E_1E_2^5\rangle$	0	0	0	-1	2	-4	$-5N_2$	-1	$-6N_2$	$-(5N-12)N_2$	$-N_2$
$\langle E_1^2E_2^4\rangle$	0	1	0	N_2	$4N-7$	N_1	$(6N-11)N_2$	1	$(N+3)N_2$	$(2N^2-7N+4)N_2$	$-N_2$
$\langle E_1^3E_2^3\rangle$	0	0	0	0	-3	0	$-6N_2$	-1	$-6N_2$	$-5N_2N_3$	$2N_2$
$\langle E_1E_2E_3^4\rangle$	0	0	0	-1	0	0	$-(4N-6)$	0	$-N_3$	$-2N_2N_4$	2
$\langle E_1E_2^2E_3^3\rangle$	0	0	0	0	-1	-1	$-(3N-10)$	0	$-N_3$	$-(2N^2-12N+19)$	-1
$\langle E_1^2E_2^2E_3^2\rangle$	0	0	1	0	$3N_3$	3	$3N_2N_3$	0	$3N_3$	$N^3-6N^2+11N-4$	2
$\langle E_1E_2E_3E_4^3\rangle$	0	0	0	0	0	0	3	0	3	$6N_3$	0
$\langle E_1E_2E_3^2E_4^2\rangle$	0	0	0	0	-1	0	$-2N_4$	0	-1	$-(N^2-7N+14)$	0
$\langle E_1E_2E_3E_4E_5^2\rangle$	0	0	0	0	0	0	3	0	0	$3N_5$	0
$\langle E_1E_2E_3E_4E_5E_6\rangle$	0	0	0	0	0	0	0	0	0	0	-15

TABLE IV. Eigenvalue moments, with $n = 2$, expressed in terms of averages of matrix elements.

	$\langle x^2 \rangle$	$\langle \lvert y \rvert^2 \rangle$
$\langle E_1^2 \rangle$	1	N_1
$\langle E_1 E_2 \rangle$	0	-1

(28) (with the appropriate choices of the n_i), (31), and (32). The results are given in Tables IV–VI. These tables are to be read in the same manner as Tables I–III. For convenience the symbol

$$N_i \equiv N - i, \quad i = 1, 2, 3, \cdots, \tag{39}$$

has been introduced in these tables.

5. A PARTICULAR CASE

For definiteness we now consider the Gaussian ensemble

$$p(H) = Ce^{-\alpha \operatorname{Tr} R^2} \exp\left(-2\gamma \sum_{i>j} S_{ij}^2\right), \tag{40}$$

TABLE V. Eigenvalue moments, with $n = 4$, expressed in terms of averages of matrix elements.

	$\langle x^4 \rangle$	$\{\langle x^2 \rangle\}^2$	$\langle x^2 \rangle \langle \lvert y \rvert^2 \rangle$	$\langle \lvert y \rvert^4 \rangle$	$\{\langle \lvert y \rvert^2 \rangle\}^2$
$\langle E_1^4 \rangle$	1	0	$4N_1$	N_1	$2N_2N_1$
$\langle E_1 E_2^3 \rangle$	0	0	-1	-1	$-2N_2$
$\langle E_1^2 E_2^2 \rangle$	0	1	$2N_2$	1	N_2N_1
$\langle E_1 E_2 E_3^2 \rangle$	0	0	-1	0	$-N_3$
$\langle E_1 E_2 E_3 E_4 \rangle$	0	0	0	0	3

where

$$C = \{2^{N(N-1)/2} \alpha^{N(N+1)/4} \gamma^{N(N-1)/4}\} / \pi^{N^2/2}. \tag{41}$$

This is the ensemble which has received the most attention in connection with the problem mentioned above of a system with a small time-reversal odd term in the Hamiltonian.[7–12,22]

It is easily verified for this distribution that

$$\langle x^{2n} \rangle = (2n-1)!!/2^n \alpha^n, \quad n = 1, 2, 3, \cdots, \tag{42}$$

$$\langle \lvert y \rvert^2 \rangle = 1/4\alpha + 1/4\gamma, \tag{43}$$

$$\langle y^2 \rangle = 1/4\alpha - 1/4\gamma, \tag{44}$$

$$\langle \lvert y \rvert^4 \rangle = 3/16\alpha^2 + 2/16\alpha\gamma + 3/16\gamma^2, \tag{45}$$

and

$$\langle \lvert y \rvert^6 \rangle = 15/64\alpha^3 + 9/64\alpha\gamma^2 + 9/64\alpha^2\gamma + 15/64\gamma^3. \tag{46}$$

The explicit expressions for the eigenvalue moments with $n = 2, 4, 6$ can now be easily obtained by inserting these results into the expressions given in Tables IV–VI. The resulting expressions are given in Table VII.

There are three interesting limiting cases which we now consider. If $\gamma \to +\infty$, the ensemble given by (40) goes to the orthogonal Gaussian ensemble.[23] If $\alpha = \gamma$, the ensemble becomes the unitary Gaussian ensemble. The results for these two special cases are given in Table VIII.

TABLE VII. Eigenvalue moments for the Gaussian ensemble.

$$\langle E_1^2 \rangle = (1/4\alpha)\,[(N+1) + (N-1)\epsilon]$$
$$\langle E_1 E_2 \rangle = -(1/4\alpha)\,[1 + \epsilon]$$
$$\langle E_1^4 \rangle = (1/16\alpha^2)\,[(2N^2 + 5N + 5) + 2(2N^2 - N - 1)\epsilon + (2N^2 - 3N + 1)\epsilon^2]$$
$$\langle E_1 E_2^3 \rangle = -(1/16\alpha^2)\,[(2N+1) + 4(N-1)\epsilon + (2N-1)\epsilon^2]$$
$$\langle E_1^2 E_2^2 \rangle = (1/16\alpha^2)\,[(N^2 + N + 1) + 2(N^2 - N - 1)\epsilon + (N^2 - 3N + 5)\epsilon^2]$$
$$\langle E_1 E_2 E_3^2 \rangle = -(1/16\alpha^2)\,[(N-1) + 2(N-2)\epsilon + (N-3)\epsilon^2]$$
$$\langle E_1 E_2 E_3 E_4 \rangle = (3/16\alpha^2)\,[1 + 2\epsilon + \epsilon^2]$$
$$\langle E_1^6 \rangle = (1/64\alpha^3)\,[(5N^3 + 22N^2 + 52N + 41) + (15N^3 + 6N^2 + 6N - 27)\epsilon + (15N^3 - 18N^2 + 12N - 9)\epsilon^2 + (5N^3 - 10N^2 + 10N - 5)\epsilon^3]$$
$$\langle E_1^2 E_2^4 \rangle = (1/64\alpha^3)\,[(2N^3 + 4N^2 + 8N + 1) + (6N^3 - 4N^2 + 2N - 19)\epsilon + (6N^3 - 16N^2 + 16N - 1)\epsilon^2 + (2N^3 - 8N^2 + 22N - 13)\epsilon^3]$$
$$\langle E_1 E_2^5 \rangle = -(1/64\alpha^3)\,[(5N^2 + 7N + 9) + (15N^2 - 19N + 7)\epsilon + (15N^2 - 23N + 19)\epsilon^2 + (5N^2 - 5N + 5)\epsilon^3]$$
$$\langle E_1^3 E_2^3 \rangle = -(1/64\alpha^3)\,[(5N^2 + 3N + 1) + (15N^2 - 15N - 9)\epsilon + (15N^2 - 39N + 27)\epsilon^2 + (5N^2 - 5N + 5)\epsilon^3]$$
$$\langle E_1 E_2^3 E_3^2 \rangle = -(1/64\alpha^3)\,[(2N^2 - 3N + 1) + (6N^2 - 19N + 7)\epsilon + (6N^2 - 25N + 31)\epsilon^2 + (2N^2 - 9N + 9)\epsilon^3]$$
$$\langle E_1 E_2^4 E_3 \rangle = -(1/64\alpha^3)\,[(2N^2 - N + 5) + (6N^2 - 15N + 27)\epsilon + (6N^2 - 23N + 15)\epsilon^2 + (2N^2 - 9N + 9)\epsilon^3]$$
$$\langle E_1 E_2 E_3^2 E_4^2 \rangle = -(1/64\alpha^3)\,[(N^2 - 3N + 5) + (3N^2 - 13N + 19)\epsilon + (3N^2 - 17N + 31)\epsilon^2 + (N^2 - 7N + 17)\epsilon^3]$$
$$\langle E_1^2 E_2^2 E_3^2 \rangle = (1/64\alpha^3)\,[(N^3 + 2N - 3) + (3N^3 - 6N^2 - 15)\epsilon + (3N^3 - 12N^2 + 18N + 3)\epsilon^2 + (N^3 - 6N^2 + 20N - 33)\epsilon^3]$$
$$\langle E_1 E_2 E_3 E_4^3 \rangle = (3/64\alpha^3)\,[(2N-1) + (6N-9)\epsilon + (6N-11)\epsilon^2 + (2N-3)\epsilon^3]$$
$$\langle E_1 E_2 E_3 E_4 E_5^2 \rangle = (3/64\alpha^3)\,[(N-3) + (3N-11)\epsilon + (3N-13)\epsilon^2 + (N-5)\epsilon^3]$$
$$\langle E_1 E_2 E_3 E_4 E_5 E_6 \rangle = -(15/64\alpha^3)\,[1 + 3\epsilon + 3\epsilon^2 + \epsilon^3]$$

TABLE VIII. Eigenvalue moments for the orthogonal and unitary Gaussian ensembles.

	Orthogonal ($\gamma \to +\infty$)	Unitary ($\alpha = \gamma$)
$\langle E_1^2 \rangle$	$\frac{N+1}{4\alpha}$	$\frac{N}{2\alpha}$
$\langle E_1 E_2 \rangle$	$-\frac{1}{4\alpha}$	$-\frac{1}{2\alpha}$
$\langle E_1^4 \rangle$	$\frac{2N^2+5N+5}{16\alpha^2}$	$\frac{2N^2+1}{4\alpha^2}$
$\langle E_1 E_2^3 \rangle$	$-\frac{2N+1}{16\alpha^2}$	$-\frac{2N-1}{4\alpha^2}$
$\langle E_1^2 E_2^2 \rangle$	$\frac{N^2+N+1}{16\alpha^2}$	$\frac{N^2-N+1}{4\alpha^2}$
$\langle E_1 E_2 E_3^2 \rangle$	$-\frac{N-1}{16\alpha^2}$	$-\frac{N-2}{4\alpha^2}$
$\langle E_1 E_2 E_3 E_4 \rangle$	$\frac{3}{16\alpha^2}$	$\frac{3}{4\alpha^2}$
$\langle E_1^6 \rangle$	$\frac{5N^3+22N^2+52N+41}{64\alpha^3}$	$\frac{5N^2(N+2)}{8\alpha^3}$
$\langle E_1^2 E_2^4 \rangle$	$\frac{2N^3+4N^2+8N+1}{64\alpha^3}$	$\frac{2N^3-3N^2+6N-4}{8\alpha^3}$
$\langle E_1 E_2^5 \rangle$	$-\frac{5N^2+7N+9}{64\alpha^3}$	$-\frac{5(N^2-N-1)}{8\alpha^3}$
$\langle E_1^3 E_2^3 \rangle$	$-\frac{5N^2+3N+1}{64\alpha^3}$	$-\frac{5N^2-7N+3}{8\alpha^3}$
$\langle E_1 E_2^2 E_3^3 \rangle$	$-\frac{2N^2-3N+1}{64\alpha^3}$	$-\frac{2N^2-7N+6}{8\alpha^3}$
$\langle E_1 E_2 E_3^4 \rangle$	$-\frac{2N^2-N+5}{64\alpha^3}$	$-\frac{2N^2-6N+7}{8\alpha^3}$
$\langle E_1 E_2 E_3^2 E_4^2 \rangle$	$-\frac{N^2-3N+5}{64\alpha^3}$	$-\frac{N^2-5N+9}{8\alpha^3}$
$\langle E_1^2 E_2^2 E_3^2 \rangle$	$\frac{N^3+2N-3}{64\alpha^3}$	$\frac{N^3-3N^2+5N-6}{8\alpha^3}$
$\langle E_1 E_2 E_3 E_4^3 \rangle$	$\frac{3(2N-1)}{64\alpha^3}$	$\frac{3(2N-3)}{8\alpha^3}$
$\langle E_1 E_2 E_3 E_4 E_5^2 \rangle$	$\frac{3(N-3)}{64\alpha^3}$	$\frac{3(N-4)}{8\alpha^3}$
$\langle E_1 E_2 E_3 E_4 E_5 E_6 \rangle$	$-\frac{15}{64\alpha^3}$	$-\frac{15}{8\alpha^3}$

Finally, if $N \to +\infty$ it can be seen from Table VII that

$$\langle E_1^{n_1} E_2^{n_2} \cdots E_6^{n_6} \rangle \sim \langle E_1^{n_1} \cdots E_6^{n_6} \rangle_{\text{orth}} (1+\epsilon)^{n/2}, \quad n = 2, 4, 6, \quad (47)$$

where $\langle E_1^{n_1} \cdots E_6^{n_6} \rangle_{\text{orth}}$ is the average obtained using the orthogonal ensemble with $N \to +\infty$, and

$$\epsilon \equiv \alpha/\gamma. \quad (48)$$

One is tempted to conjecture at this point that (47) is true for all positive values of n.[24] If this is true, it would appear that the relative size of α and γ affects the eigenvalue moments only with respect to scale. Further, since $\epsilon = 0$ and $\epsilon = 1$ correspond to the orthogonal and unitary ensembles, respectively, it would then follow that

$$\langle E_1^{n_1} \cdots E_m^{n_m} \rangle_{\text{unitary}} \sim 2^{n/2} \langle E_1^{n_1} \cdots E_m^{n_m} \rangle_{\text{orth}} \quad (49)$$

for large N.

6. SUMMARY AND DISCUSSION

In this article we have shown that the evaluation of averages of the form $\langle E_1^{n_1} \cdots E_m^{n_m} \rangle$, where the n_i are nonnegative integers, can always be reduced to evaluation of averages involving products of matrix elements. Thus, such averages can be evaluated without finding the joint eigenvalue distribution explicitly.

There are at least two possible applications where knowledge of these averages could be extremely useful. First, if an approximate expression for any n-level correlation function is found for some matrix element distribution, these averages could be used to check the validity of the approximation.[25]

A second possible application may be in the calculation of the moments of the distribution of widths.

To date there has been little success in obtaining the distribution of widths or even its moments for the ensemble given by (40). Of particular interest are the moments in the limiting case $\alpha \ll \gamma$ and $N \to +\infty$. It seems to us that the eigenvalue moments may be crucial in obtaining these moments for the distribution of widths.

* Supported in part by an N.R.C. grant.

[1] See, for example, *Statistical Theories of Spectra: Fluctuations*, C. E. Porter, Ed. (Academic, New York, 1965). Many of the pertinent papers (including Refs. 2–7 below) are contained in this collection, as well as an excellent introductory review of this subject. See also M. L. Mehta, *Random Matrices and the Statistical Theory of Energy Levels* (Academic, New York, 1967).

[2] C. E. Porter and R. G. Thomas, Phys. Rev. **104**, 483 (1956).

[3] N. Ullah, J. Math. Phys. **4**, 1279 (1963).

[4] C. E. Porter and N. Rosenzweig, Ann. Acad. Sci. Fenn., Ser. A6: Phys., No. 44 (1960).

[5] F. J. Dyson, J. Math. Phys. **3**, 140 (1962).

[6] N. Rosenzweig and C. E. Porter, Phys. Rev. **120**, 1698 (1960).

[7] N. Rosenzweig, J. E. Monahan, and M. L. Mehta, Nucl. Phys. **A109**, 437 (1968).

[8] M. L. Mehta and N. Rosenzweig, Nucl. Phys. **A109**, 449 (1968).

[9] M. L. Mehta, Nuovo Cimento **65B**, 107 (1970).

[10] L. D. Favro and J. F. McDonald, Phys. Rev. Letters **19**, 1254 (1967).

[11] L. D. Favro and J. F. McDonald, J. Math. Phys. **9**, 1429 (1968).

[12] J. F. McDonald, J. Math. Phys. **10**, 1191 (1969).

[13] J. F. McDonald and L. D. Favro, J. Math. Phys. **11**, 3103 (1970).

[14] The results in Ref. 8 are an exception. However, the case considered there seems to have little physical interest.

[15] F. J. Dyson, J. Math. Phys. **3**, 1199 (1962).

[16] As previously noted, a permutation of indices on the eigenvalues does not change the average of a quantity. Thus, by a distinct term we mean one that cannot be obtained from another by a simple permutation of the indices. For example, if $m = 2$,

$$\sum_{\alpha=1}^{2} \langle E_1^{n_\alpha} E_2^{n-n_\alpha} \rangle = \langle E_1^{n_1} E_2^{n_2} \rangle + \langle E_1^{n} E_2^{n_1} \rangle.$$

However, since these terms are identical, the second is not distinct and thus not to be included in $\sum_{\alpha=1}^{2}{}'$.

[17] Of course, for a particular set of n_α the averages required include only those for which each s_α is a sum of the n_α, each n_α occurring once and only once. For example, to evaluate $\langle E_1 E_2^2 E_3^3 \rangle$, one needs only $\langle E_1^3 E_2^3 \rangle$, $\langle E_1^2 E_2^1 \rangle$, $\langle E_1 E_2^5 \rangle$, and $\langle E_1^6 \rangle$.

[18] One might still question the second statement if one or more of the n_i are zero. However, (29) is still valid in this case provided that $E_i^0 \equiv 1$ and $\mathrm{Tr}\, H^0 \equiv N$.

[19] These can also be established by noting that coefficients, in fact, can be found by a counting procedure to be discussed below.

[20] By definition two terms are of the same type if one can be obtained from the other by a permutation of the n_i.

[21] Another approach (basically equivalent to the one given above) which could be used is to express the $\langle \mathrm{Tr}\, H^{s_1} \cdots \mathrm{Tr}\, H^{s_\alpha} \rangle$ occurring in (29) in terms of averages of eigenvalues. This will give a system of simultaneous equations for the moments of the eigenvalues. Then $\langle E_1^{n_1} \cdots E_\alpha^{n_\alpha} \rangle$, $\alpha = 1, \cdots, m$, for arbitrary n_i could be determined by solving these equations.

[22] It should be noted that for this particular ensemble various useful identities can be found. For example,

$$\langle Q(E)\, \mathrm{Tr}\, H^2 \rangle = -\left(\frac{\partial}{\partial \alpha} + \frac{\partial}{\partial \gamma'}\right)\langle Q(E) \rangle + \left(\frac{N(N+1)}{4\alpha} + \frac{N(N-1)}{4\gamma'}\right)\langle Q(E) \rangle$$

and

$$N\langle Q(E) \rangle + \sum_{l=1}^{N} \left\langle \mathrm{Tr}\, H \frac{\partial Q(E)}{\partial E_l} \right\rangle = 2\alpha \langle (\mathrm{Tr}\, H)^2 Q(E) \rangle.$$

[23] Some of the results given in Table VIII were previously obtained via other methods. See M. L. Mehta, Nucl. Phys. **18**, 395 (1960); N. Ullah and C. E. Porter, Phys. Letters **6**, 301 (1963); and N. Ullah, Nucl. Phys. **58**, 65 (1964). In the last article Ullah conjectured that for the Gaussian orthogonal ensemble

$$\langle E_1 E_2 \cdots E_{2m} \rangle = (-1)^m (2m-1)!!/(4\alpha)^m.$$

The results in Tables IV–VI suggest that this conjecture could be generalized to

$$\langle E_1 E_2 \cdots E_{2m} \rangle = (-1)^m (2m-1)!! \{\langle |y|^2 \rangle\}^m,$$

for any ensemble.

[24] It can be shown that if (47) is valid for $\langle E_1^{n_1} \cdots E_m^{n_m} \rangle$, then it is also true for $\langle E_1^{n_1} \cdots E_m^{n_m} E_{m+1}^2 \rangle$.

[25] The results given in Refs. 7 and 9 above are not suitable for such a check.

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 12, NUMBER 3 MARCH 1971

Integral Representation Technique for Expansion of Arbitrary Analytic Functions of the Distance between Two Points

A. K. RAFIQULLAH
Department of Physics, University of Dacca, Dacca, Pakistan

(Received 11 September 1970)

The coefficients of expansion of an arbitrary analytic function of the distance r between two points (r_1, θ_1, ϕ_1) and (r_2, θ_2, ϕ_2) in terms of the Legendre polynomials $P_l(\cos\theta_{12})$ are double Bessel transformed. Assuming that the transformed coefficients are diagonal, consistent with the differential equations satisfied by the original coefficients, we derive the explicit expressions for the latter coefficients. These formulas are identical to those derived by Sack from the solutions of the differential equations in terms of the hypergeometric functions.

1. INTRODUCTION

The expansion of the inverse distance r^{-1} between two points 1 and 2, with spherical coordinates (r_1, θ_1, ϕ_1) and (r_2, θ_2, ϕ_2) in terms of the Legendre polynomials $P_l(\cos\theta_{12})$, is given by the Laplace expansion

$$r^{-1} = r_>^{-1} \sum_{l=0}^{\infty} \left(\frac{r_<}{r_>}\right)^l P_l(\cos\theta_{12}), \tag{1}$$

where

$$r_< = \min(r_1, r_2), \quad r_> = \max(r_1, r_2), \tag{2}$$

$$\cos\theta_{12} = \cos\theta_1 \cos\theta_2 + \sin\theta_1 \sin\theta_2 \cos(\phi_1 - \phi_2). \tag{3}$$

More frequently, we encounter arbitrary powers of r and require expansions in terms of Legendre polynomials as in (1),

$$r^n = \sum_l R_{nl}(r_1, r_2) P_l(\cos\theta_{12}), \tag{4}$$

where the radial function R_{nl} satisfies the differential equation

$$\frac{\partial^2 R_{nl}}{\partial r_1^2} + \frac{2}{r_1}\frac{\partial R_{nl}}{\partial r_1} - \frac{l(l+1)R_{nl}}{r_1^2} = \frac{\partial^2 R_{nl}}{\partial r_2^2} + \frac{2}{r_2}\frac{\partial R_{nl}}{\partial r_2} - \frac{l(l+1)R_{nl}}{r_2^2}. \tag{5}$$

The expansion (4) for $n = 1$ has been given explicitly by Jen.[1] The cases for $n = -1$ and -2 have been treated by Fontana[2] by group-theoretical methods. The formulas for the general case have been derived

[16] As previously noted, a permutation of indices on the eigenvalues does not change the average of a quantity. Thus, by a distinct term we mean one that cannot be obtained from another by a simple permutation of the indices. For example, if $m = 2$,

$$\sum_{\alpha=1}^{2} \langle E_1^{n_\alpha} E_2^{n-n_\alpha} \rangle = \langle E_1^{n_1} E_2^{n_2} \rangle + \langle E_1^{n} E_2^{n_1} \rangle.$$

However, since these terms are identical, the second is not distinct and thus not to be included in $\sum_{\alpha=1}^{2}{}'$.

[17] Of course, for a particular set of n_α the averages required include only those for which each s_α is a sum of the n_α, each n_α occurring once and only once. For example, to evaluate $\langle E_1 E_2^2 E_3^3 \rangle$, one needs only $\langle E_1^3 E_2^3 \rangle$, $\langle E_1^2 E_2^1 \rangle$, $\langle E_1 E_2^5 \rangle$, and $\langle E_1^6 \rangle$.

[18] One might still question the second statement if one or more of the n_i are zero. However, (29) is still valid in this case provided that $E_i^0 \equiv 1$ and $\mathrm{Tr}\, H^0 \equiv N$.

[19] These can also be established by noting that coefficients, in fact, can be found by a counting procedure to be discussed below.

[20] By definition two terms are of the same type if one can be obtained from the other by a permutation of the n_i.

[21] Another approach (basically equivalent to the one given above) which could be used is to express the $\langle \mathrm{Tr}\, H^{s_1} \cdots \mathrm{Tr}\, H^{s_\alpha} \rangle$ occurring in (29) in terms of averages of eigenvalues. This will give a system of simultaneous equations for the moments of the eigenvalues. Then $\langle E_1^{n_1} \cdots E_\alpha^{n_\alpha} \rangle$, $\alpha = 1, \cdots, m$, for arbitrary n_i could be determined by solving these equations.

[22] It should be noted that for this particular ensemble various useful identities can be found. For example,

$$\langle Q(E)\, \mathrm{Tr}\, H^2 \rangle = -\left(\frac{\partial}{\partial \alpha} + \frac{\partial}{\partial \gamma'}\right)\langle Q(E) \rangle + \left(\frac{N(N+1)}{4\alpha} + \frac{N(N-1)}{4\gamma'}\right)\langle Q(E) \rangle$$

and

$$N\langle Q(E) \rangle + \sum_{l=1}^{N} \left\langle \mathrm{Tr}\, H \frac{\partial Q(E)}{\partial E_l} \right\rangle = 2\alpha \langle (\mathrm{Tr}\, H)^2 Q(E) \rangle.$$

[23] Some of the results given in Table VIII were previously obtained via other methods. See M. L. Mehta, Nucl. Phys. **18**, 395 (1960); N. Ullah and C. E. Porter, Phys. Letters **6**, 301 (1963); and N. Ullah, Nucl. Phys. **58**, 65 (1964). In the last article Ullah conjectured that for the Gaussian orthogonal ensemble

$$\langle E_1 E_2 \cdots E_{2m} \rangle = (-1)^m (2m-1)!!/(4\alpha)^m.$$

The results in Tables IV–VI suggest that this conjecture could be generalized to

$$\langle E_1 E_2 \cdots E_{2m} \rangle = (-1)^m (2m-1)!! \{\langle |y|^2 \rangle\}^m,$$

for any ensemble.

[24] It can be shown that if (47) is valid for $\langle E_1^{n_1} \cdots E_m^{n_m} \rangle$, then it is also true for $\langle E_1^{n_1} \cdots E_m^{n_m} E_{m+1}^2 \rangle$.

[25] The results given in Refs. 7 and 9 above are not suitable for such a check.

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 12, NUMBER 3 MARCH 1971

Integral Representation Technique for Expansion of Arbitrary Analytic Functions of the Distance between Two Points

A. K. RAFIQULLAH
Department of Physics, University of Dacca, Dacca, Pakistan

(Received 11 September 1970)

The coefficients of expansion of an arbitrary analytic function of the distance r between two points (r_1, θ_1, ϕ_1) and (r_2, θ_2, ϕ_2) in terms of the Legendre polynomials $P_l(\cos\theta_{12})$ are double Bessel transformed. Assuming that the transformed coefficients are diagonal, consistent with the differential equations satisfied by the original coefficients, we derive the explicit expressions for the latter coefficients. These formulas are identical to those derived by Sack from the solutions of the differential equations in terms of the hypergeometric functions.

1. INTRODUCTION

The expansion of the inverse distance r^{-1} between two points 1 and 2, with spherical coordinates (r_1, θ_1, ϕ_1) and (r_2, θ_2, ϕ_2) in terms of the Legendre polynomials $P_l(\cos\theta_{12})$, is given by the Laplace expansion

$$r^{-1} = r_>^{-1} \sum_{l=0}^{\infty} \left(\frac{r_<}{r_>}\right)^l P_l(\cos\theta_{12}), \tag{1}$$

where

$$r_< = \min(r_1, r_2), \quad r_> = \max(r_1, r_2), \tag{2}$$

$$\cos\theta_{12} = \cos\theta_1 \cos\theta_2 + \sin\theta_1 \sin\theta_2 \cos(\phi_1 - \phi_2). \tag{3}$$

More frequently, we encounter arbitrary powers of r and require expansions in terms of Legendre polynomials as in (1),

$$r^n = \sum_l R_{nl}(r_1, r_2) P_l(\cos\theta_{12}), \tag{4}$$

where the radial function R_{nl} satisfies the differential equation

$$\frac{\partial^2 R_{nl}}{\partial r_1^2} + \frac{2}{r_1}\frac{\partial R_{nl}}{\partial r_1} - \frac{l(l+1)R_{nl}}{r_1^2} = \frac{\partial^2 R_{nl}}{\partial r_2^2} + \frac{2}{r_2}\frac{\partial R_{nl}}{\partial r_2} - \frac{l(l+1)R_{nl}}{r_2^2}. \tag{5}$$

The expansion (4) for $n = 1$ has been given explicitly by Jen.[1] The cases for $n = -1$ and -2 have been treated by Fontana[2] by group-theoretical methods. The formulas for the general case have been derived

by Sack[3] from the solutions of (5) in terms of hypergeometric functions satisfying continuity and dimensionality conditions. The transformation theory of hypergeometric functions have been extensively applied by him to express R_{nl} in various forms, some of which resulted in forms symmetric in $r_<$ and $r_>$. By generalizing (4) to arbitrary analytic function of r, namely,

$$f(r) = \sum_l f_l(r_<, r_>)P_l(\cos\theta_{12}), \tag{6}$$

Sack was also able to give three analogous formulas for f_l involving spherical Bessel functions of $(r_<\partial/\partial r_>)$ of related operators acting on $f(r_>)$, $f(r_1 + r_2)$, or $f[(r_1^2 + r_2^2)^{\frac{1}{2}}]$.

Since the simplicity and elegance of integral transforms (and representations) sometimes causes this technique to be adopted, rather than trying to solve the corresponding untidy differential equations, it is the purpose of the present paper to derive explicitly the formulas for the coefficients f_l in the expansion (6) with the aid of such a technique. It is in this respect that our approach differs from that of Sack. In Sec. 2 we give the formalism. The formulas thus obtained are identical to those given by Sack.[3] Finally, we quote the results of application of the formulas to some functions of physical interest.

2. FORMALISM

The coefficients $f_l(r_<, r_>)$ in (6) are double Bessel transformed:

$$f_l(r_<, r_>) = \int_0^\infty k_>^2\, dk_> \int_0^\infty k_<^2\, dk_< j_l(k_<r_<) \times F_l(k_<, k_>)j_l(k_>r_>), \tag{7}$$

where $F_l(k_<, k_>)$ is the transform of $f_l(r_<, r_>)$ and is given by the inversion formula

$$F_l(k_<, k_>) = \int_0^\infty r_>^2\, dr_> \int_0^\infty r_<^2\, dr_< j_l(k_<r_<) \times f_l(r_<, r_>)j_l(k_>r_>). \tag{8}$$

Expression (7) must satisfy the radial equation (5). In order that $F_l(k_<, k_>)$ be simpler to work with rather than $f_l(r_<, r_>)$, it must have a diagonal form, that is, it must be proportional to the delta function $\delta(k_> - k_<)$. In fact, we see that if we choose the diagonal form

$$F_l(k_<, k_>) = (2l + 1)F(k_>)k_>^{-2}\delta(k_> - k_<), \tag{9}$$

then

$$f_l(r_<, r_>) = (2l + 1)\int_0^\infty k^2\, dk\, j_l(kr_<)F(k)j_l(kr_>), \tag{10}$$

where we have set $k_> = k$, satisfies (5). The factor $(2l + 1)$ has been incorporated so that application of (10) to a function whose expansion is known, say Laplace expansion (1), gives the desired expansion.

By letting $r_< = 0$, it follows immediately from (6) and (10) that $F(k)$ is the Bessel transform of

$$f(r_>) = \int_0^\infty k^2\, dk\, j_0(kr_>)F(k). \tag{11}$$

Equation (10) is our working formula. In order to evaluate the integral, we expand $j_l(kr_<)$ in power series, so that

$$f_l(r_<, r_>) = (2l + 1)\sum_{s=0}^\infty \frac{(-)^s r_<^{2s+l}}{(2s)!!\,(2s + 2l + 1)!!} \times \int_0^\infty F(k)j_l(kr_>)k^{2s+l+2}\, dk. \tag{12}$$

Since

$$r^{-1}\left(\frac{\partial}{\partial r}\right)^{2s} rj_0(kr) = (-)^s k^{2s} j_0(kr) \tag{13}$$

and

$$(-r)^l\left(r^{-1}\frac{\partial}{\partial r}\right)^l j_0(kr) = k^l j_l(kr), \tag{14}$$

we find, with the help of (11),

$$(-r_>)^l\left(r_>^{-1}\frac{\partial}{\partial r_>}\right)^l r_>^{-1}\left(\frac{\partial}{\partial r_>}\right)^{2s} r_>f(r_>) = (-)^s\int_0^\infty F(k)j_l(kr_>)k^{2s+l+2}\, dk. \tag{15}$$

Substitution of (15) into the right side of (12) yields

$$f_l(r_<, r_>) = (2l + 1)\sum_{s=0}^\infty \frac{(-r_<r_>)^l r_<^{2s}}{(2s)!!\,(2s + 2l + 1)!!} \times \left(r_>^{-1}\frac{\partial}{\partial r_>}\right)^l\left[r_>^{-1}\left(\frac{\partial}{\partial r_>}\right)^{2s} r_>f(r_>)\right], \tag{16a}$$

or operational expansion yields

$$f_l(r_<, r_>) = (2l + 1)(-r_<r_>)^l(r_>^{-1}\partial/\partial r_>)^l \times \left(r_>^{-1}\frac{i_l(r_<\partial/\partial r_>)}{(r_<\partial/\partial r_>)^l}[r_>f(r_>)]\right), \tag{16b}$$

where i_l is the modified Bessel function

$$i_l(z) = \sum_{s=0}^\infty \frac{z^{l+2s}}{(2s)!!\,(2s + 2l + 1)!!}. \tag{17}$$

Writing (10) in the form

$$f_l(r_1, r_2) = (2l + 1)\int_0^\infty j_l(kr_1)F(k)j_l(kr_2)k^2\, dk \tag{18}$$

and in the integrand using the following formula,

with $r_+ = r_1 + r_2$,

$$i_l(kr_1)j_l(kr_2) = \sum_{s=0}^{\infty} \frac{2^s(l+1)_s}{(2l+1)!!\,(2l+2)_s s!}\left(\frac{kr_1r_2}{r_+}\right)^{l+s} j_{l+s}(kr_+), \quad (19)$$

where

$$(a)_0 = 1, \quad (a)_s = a(a+1)\cdots(a+s-1) = \Gamma(a+s)/\Gamma(a), \quad (20)$$

we obtain

$$f_l(r_1, r_2) = (2l+1)\sum_{s=0}^{\infty} \frac{2^s(l+1)_s}{(2l+1)!!\,(2l+2)_s s!}\left(\frac{r_1r_2}{r_+}\right)^{l+s} \times \int_0^{\infty} F(k) j_{l+s}(kr_+) k^{l+s+2}\, dk. \quad (21)$$

Since it follows from (11) that

$$f(r_+) = \int_0^{\infty} F(k) j_0(kr_+) k^2\, dk, \quad (22)$$

we can evaluate the integral in (21) with the help of (22). Thus

$$f_l = \frac{1}{(2l-1)!!}\left(\frac{-r_1r_2}{r_+}\frac{\partial}{\partial r_+}\right)^l \times \sum_{s=0}^{\infty} \frac{(l+1)_s}{(2l+2)_s s!}\left(\frac{-2r_1r_2}{r_+}\frac{\partial}{\partial r_+}\right)^s f(r_+)$$

$$= \frac{1}{(2l-1)!!}\left(\frac{-r_1r_2}{r_+}\frac{\partial}{\partial r_+}\right)^l \times \Phi\left(l+1; 2l+2; \frac{-2r_1r_2}{r_+}\frac{\partial}{\partial r_+}\right) f(r_+), \quad (23)$$

where Φ is the confluent hypergeometric function and where the product r_1r_2 is to be treated as constant in the differentiation. Now, using

$$i_l(z) = \frac{1}{(2l+1)!!} z^l e^{-z} \Phi(l+1; 2l+2; 2z) \quad (24)$$

and, for $\mathrm{Re}\, a > 0$, $\mathrm{Re}\,(b-a) > 0$,

$$\Phi(a; b; z) = e^z \Phi(b-a; b; -z), \quad (25)$$

we can write (23) as

$$f_l(r_1, r_2) = (2l+1) i_l\left(\frac{-r_1r_2}{r_+}\frac{\partial}{\partial r_+}\right) \times \exp\left(\frac{-r_1r_2}{r_+}\frac{\partial}{\partial r_+}\right) f(r_+). \quad (26)$$

Finally, using the Taylor expansion

$$\exp\left(h\frac{\partial}{\partial z}\right) f(z) = f(z+h) \quad (27)$$

and

$$\left(z^{-1}\frac{\partial}{\partial z}\right) = 2\frac{\partial}{\partial (z^2)}, \quad (28)$$

introducing the variable $\rho = (r_1^2 + r_2^2)^{\frac{1}{2}}$, and noting that the product r_1r_2 is to be treated as constant in the differentiation, we can write (26) operationally as

$$f_l(r_1, r_2) = (2l+1) i_l\left(\frac{-r_1r_2}{r_+}\frac{\partial}{\partial r_+}\right) f[(r_+^2 - 2r_1r_2)^{\frac{1}{2}}] = (2l+1) i_l\left(\frac{-r_1r_2}{\rho}\frac{\partial}{\partial \rho}\right) f(\rho). \quad (29)$$

Formulas (16), (23), and (29) are identical to those derived by Sack[3] on the basis of solving the differential equation satisfied by the radial function f_l in terms of hypergeometric functions.

We give below the results of the application of the above formulas to some functions of physical interest.

(i) Let $f(r) = w_0(kr)$, where $r = |\mathbf{r}_1 - \mathbf{r}_2|$ and w_0 is a Bessel function,

$$w_0(kr) = j_0(kr), \quad n_0(kr), \quad h_0^{(1)}, \quad h_0^{(2)}(kr). \quad (30)$$

Application of (16) gives

$$w_0(kr) = \sum_l (2l+1) j_l(kr_<) w_l(kr_>) P_l(\cos\theta_{12}). \quad (31)$$

(ii) If $f(r)$ is a modified Bessel function $i_0(kr)$ or $k_0(kr)$, then

$$i_0(kr) = \sum_l (2l+1)(-)^l i_l(kr_<) i_l(kr_>) P_l(\cos\theta_{12}), \quad (32a)$$

$$k_0(kr) = \sum_l (2l+1) i_l(kr_<) k_l(kr_>) P_l(\cos\theta_{12}). \quad (32b)$$

(iii) If $f(r) = e^{-ar}/r$, then

$$\frac{e^{-ar}}{r} = -a\sum_l (2l+1) i_l(ar_<) i^l h_l^{(1)}(iar_>) P_l(\cos\theta_{12}). \quad (33)$$

(iv) If $f(r) = e^{ikr}/r$, then, by setting $a = -ik$ in (33), we find

$$\frac{e^{ikr}}{r} = ik\sum_l (2l+1) j_l(kr_<) h_l^{(1)}(kr_>) P_l(\cos\theta_{12}). \quad (34)$$

(v) If $f(r) = \exp(-ar^2)$, then application of (29) yields

$$\exp(-ar^2) = \sum_l (2l+1) i_l(2ar_1r_2) \times \exp[-a(r_1^2 + r_2^2)] P_l(\cos\theta_{12}). \quad (35)$$

(vi) Cancelling the common factor $\exp[-(r_1^2 + r_2^2)a]$ from both sides of (35), we find

$$\exp(2ar_1r_2\cos\theta_{12}) = \sum_l (2l+1) i_l(2ar_1r_2) P_l(\cos\theta_{12}). \quad (36)$$

(vii) If in (36) we set $2ar_1 = ik$, $r_2 = r$, and $\theta_{12} = 0$, we obtain

$$\exp(ikr\cos\theta) = \sum_l i^l(2l+1) j_l(kr) P_l(\cos\theta). \quad (37)$$

Application of (16) to $f(r) = r^{-1}$, of course, yields (1). Formulas (31), (32), (35), and (36) have also been quoted by Sack.[3]

ACKNOWLEDGMENT

The author wishes to thank Professor M. Innas Ali for the facility extended to him at the Department of Physics, University of Dacca, Pakistan.

[1] C. K. Jen, Phys. Rev. **43**, 540 (1933).
[2] P. R. Fontana, J. Math. Phys. **2**, 825 (1961).
[3] R. A. Sack, J. Math. Phys. **5**, 245 (1964).

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 12, NUMBER 3 MARCH 1971

Almost Periodic Functions and the Theory of Disordered Systems*

M. V. ROMERIO
Laboratory of Atomic and Solid State Physics, Cornell University, Ithaca, New York 14850

(Received 30 July 1970)

The Schrödinger equation appropriate to a potential described by an almost periodic function (in the sense of representation theory) is examined. The Bloch–Floquet theorem is generalized to almost periodic functions and some properties of the density of states and spatial localization of the electrons are obtained. Some qualitative aspects of band structure are discussed.

INTRODUCTION

The purpose of this paper is to propose a new approach to the problem of disordered systems, and to give some general features which can be drawn from it.

As usual, we assume that the states of the electrons can be described by the eigensolutions of a Schrödinger equation for a potential V. Essentially, this means that we restrict our discussion to systems for which the adiabatic approximation is valid.

To construct the potential V, we consider the array of ions, in which the electrons move, as a kind of oblique projection of three n-dimensional periodic lattices on three independent directions of the usual Euclidean space.

This picture is justified by the fact that a bounded potential is closely approximated by a series of unimodular exponentials—that is, by a kind of "generalized Fourier series" with several basic frequencies. As shown in the Appendix, this series, which we shall ultimately identify with V, can be obtained essentially by projection from a function W of $3n$ variables, *periodic in each variable*. W can therefore be looked at as a function generating a $3n$-dimensional lattice.

From a mathematical point of view, V is nothing more than an almost periodic function (APF) (in the sense the term is used in representation theory). A brief discussion about this wide class of functions is given in the first section. (The reader who requires more information on APF's is referred to Refs. 1–3.)

The second section is devoted to a generalization of Bloch's theorem for a periodic potential. The main idea here is to use the translational symmetry of the $3n$-dimensional lattice mentioned above and to "project" the result in three dimensions.

In Sec. 3 we use these results to discuss some features of the energy bands of the system under consideration. We show that disorder leads to the existence of a very large number of small gaps in the energy. We also notice that the concept of a Brillouin zone becomes meaningless and leads us to work in "the extended zone scheme."

The properties of the density of states and the Van Hove singularities are considered in Secs. 4 and 5. We show that, in certain cases near the edges of a band, the density of states may be characterized by a small tailing, but that Van Hove singularities give very sharp edges for the same point. (The existence of a tailing was already found in some other papers for slightly different situations. See for example Refs. 4 and 5 and references therein.) Finally, in Sec. 6, we use the Green function of a differential equation of $3n$

(vi) Cancelling the common factor $\exp[-(r_1^2 + r_2^2)a]$ from both sides of (35), we find

$$\exp(2ar_1r_2\cos\theta_{12}) = \sum_l (2l+1) i_l(2ar_1r_2) P_l(\cos\theta_{12}). \quad (36)$$

(vii) If in (36) we set $2ar_1 = ik$, $r_2 = r$, and $\theta_{12} = 0$, we obtain

$$\exp(ikr\cos\theta) = \sum_l i^l(2l+1) j_l(kr) P_l(\cos\theta). \quad (37)$$

Application of (16) to $f(r) = r^{-1}$, of course, yields (1). Formulas (31), (32), (35), and (36) have also been quoted by Sack.[3]

ACKNOWLEDGMENT

The author wishes to thank Professor M. Innas Ali for the facility extended to him at the Department of Physics, University of Dacca, Pakistan.

[1] C. K. Jen, Phys. Rev. **43**, 540 (1933).
[2] P. R. Fontana, J. Math. Phys. **2**, 825 (1961).
[3] R. A. Sack, J. Math. Phys. **5**, 245 (1964).

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 12, NUMBER 3 MARCH 1971

Almost Periodic Functions and the Theory of Disordered Systems*

M. V. ROMERIO
Laboratory of Atomic and Solid State Physics, Cornell University, Ithaca, New York 14850

(Received 30 July 1970)

The Schrödinger equation appropriate to a potential described by an almost periodic function (in the sense of representation theory) is examined. The Bloch–Floquet theorem is generalized to almost periodic functions and some properties of the density of states and spatial localization of the electrons are obtained. Some qualitative aspects of band structure are discussed.

INTRODUCTION

The purpose of this paper is to propose a new approach to the problem of disordered systems, and to give some general features which can be drawn from it.

As usual, we assume that the states of the electrons can be described by the eigensolutions of a Schrödinger equation for a potential V. Essentially, this means that we restrict our discussion to systems for which the adiabatic approximation is valid.

To construct the potential V, we consider the array of ions, in which the electrons move, as a kind of oblique projection of three n-dimensional periodic lattices on three independent directions of the usual Euclidean space.

This picture is justified by the fact that a bounded potential is closely approximated by a series of unimodular exponentials—that is, by a kind of "generalized Fourier series" with several basic frequencies. As shown in the Appendix, this series, which we shall ultimately identify with V, can be obtained essentially by projection from a function W of $3n$ variables, *periodic in each variable*. W can therefore be looked at as a function generating a $3n$-dimensional lattice.

From a mathematical point of view, V is nothing more than an almost periodic function (APF) (in the sense the term is used in representation theory). A brief discussion about this wide class of functions is given in the first section. (The reader who requires more information on APF's is referred to Refs. 1–3.)

The second section is devoted to a generalization of Bloch's theorem for a periodic potential. The main idea here is to use the translational symmetry of the $3n$-dimensional lattice mentioned above and to "project" the result in three dimensions.

In Sec. 3 we use these results to discuss some features of the energy bands of the system under consideration. We show that disorder leads to the existence of a very large number of small gaps in the energy. We also notice that the concept of a Brillouin zone becomes meaningless and leads us to work in "the extended zone scheme."

The properties of the density of states and the Van Hove singularities are considered in Secs. 4 and 5. We show that, in certain cases near the edges of a band, the density of states may be characterized by a small tailing, but that Van Hove singularities give very sharp edges for the same point. (The existence of a tailing was already found in some other papers for slightly different situations. See for example Refs. 4 and 5 and references therein.) Finally, in Sec. 6, we use the Green function of a differential equation of $3n$

variables to obtain some results on the *spatial localization of electrons* in disordered systems.

I. SOME PROPERTIES OF ALMOST PERIODIC FUNCTIONS

In this section we give (in general without proof) some important properties of APF on the group R^3 (the group associated with the addition of vectors in the usual three-dimensional space), R is the real line. Since these properties are shared with all Abelian locally compact groups, we outline the general case in the Appendix.

Definition 1: A continuous bounded function f belonging to R^3 is almost periodic if, for every $\epsilon > 0$, there exists $T > 0$ $(T \in R)$ such that every sphere (ball) of diameter T contains the extremity of a vector τ satisfying

$$\|f - f_\tau\|_\infty \leq \epsilon,$$

where f_τ is defined by $f_\tau(\mathbf{x}) = f(\mathbf{x} + \tau)$ and

$$\|f\|_\infty = \sup_{\mathbf{x} \in R^3} |f(\mathbf{x})|.$$

We show in the Appendix that such a function can be expressed as a "generalized Fourier series." More precisely, there exists a discrete subset Λ_f of R^3 such that the series

$$\sum_{\lambda \in \Lambda_f} A_\lambda e^{i(\lambda, \mathbf{x})}$$

converges uniformly towards f.

It is of interest to notice here that

$$\|f - f_{n\tau}\| \leq n\epsilon,$$

where n is an integer, which shows that APF's are certainly good candidates for the description of order at short distance.

It is clear, on the other hand, that the set of APF's is very large and is sufficient to approximate any physical bounded potential. In fact, the possibility of representing any APF as a "generalized Fourier series" shows immediately that the sums and the products of an enumerable set of *periodic* functions are also APF.

An example of such a function is given by the Bloch type solution

$$\Psi_\mathbf{k}(\mathbf{x}) = e^{i\mathbf{k}\cdot\mathbf{x}} u_\mathbf{k}(\mathbf{x}),$$

where $u_\mathbf{k}$ is a periodic function. We have, in fact, for a given ϵ

$$|\Psi_\mathbf{k}(\mathbf{x} + \tau) - \Psi_\mathbf{k}(\mathbf{x})| = |u_\mathbf{k}(\mathbf{x} + \tau)e^{i\mathbf{k}\cdot\tau} - u_\mathbf{k}(\mathbf{x})|.$$

We now choose a τ belonging to the set of periods of $u_\mathbf{k}$; then

$$|\Psi_\mathbf{k}(\mathbf{x} + \tau) - \Psi_\mathbf{k}(\mathbf{x})| \leq \sup_{\mathbf{x} \in R^3} |u_\mathbf{k}(\mathbf{x})|\,|e^{i\mathbf{k}\cdot\tau} - 1|.$$

Now τ can be chosen in such a way that $|e^{i\mathbf{k}\tau} - 1|$ is arbitrarily small. We take τ such that

$$\sup_{\mathbf{x} \in R^3} |u_\mathbf{k}(\mathbf{x})|\,|e^{i\mathbf{k}\cdot\tau} - 1| < \epsilon.$$

Definition 2: Let F be a function of $R^3 \times R^3 \cdots \times R^3$ (i.e., F is a function of several variables). We call diagonal of F (notation diag F) the function $f(\mathbf{x})$ defined by

$$f(\mathbf{x}) \equiv \text{diag } F(\mathbf{x}) = F(\mathbf{x}, \mathbf{x}, \cdots, \mathbf{x}).$$

It is possible to show (see Ref. 1, p. 35) that the definition may be extended to an application F of a direct product of an infinite number of space R^3 into C (the complex numbers).

Proposition 1: For any APF f on R^3 there exists a function F, of the type defined above, *periodic in each variable*, and such that

$$f = \text{diag } F.$$

For the proof, see Appendix.

This last proposition will be the most important result required for the establishment of the theorem giving the general form of an eigensolution of the Schrödinger equation for a potential given by an APF.

II. THE BLOCH THEOREM FOR APF's

Let $V(\mathbf{x})$ be an APF. We wish to show that the eigensolutions of the operator $-\nabla^2 + V$ on $\mathcal{K}$, the pre-Hilbert space generated by APF, are similar to those given by Bloch for a periodic potential. (A pre-Hilbert space is a space the closure of which is a Hilbert space.)

We first prove the following.

Theorem 1: Let $V(\mathbf{x})$ be an APF and W a function satisfying the conditions of Proposition 1, i.e., a periodic function such that diag $W = V$. Then there exists a linear differentiable operator D such that all the eigensolutions ϕ of $-D + W$, for which diag $\phi \neq 0$, satisfy

$$(-\nabla^2 + V(X) - E) \text{ diag } \phi = 0,$$

with ϕ belonging to the pre-Hilbert space $\mathcal{K}'$ generated by APF's on R^{3n}.

Moreover, if ψ is an eigensolution of $-\nabla^2 + V$ and $\psi \in \mathcal{K}$, there exists a function ϕ (eigensolution of $-D + W$), such that diag $\phi = \psi$.

Proof: Let $\phi \in \mathcal{K}'$. As such (see Appendix) ϕ can be expressed as a Fourier integral:

$$\phi(\mathbf{x}_1 \cdots \mathbf{x}_n) = \lim_{V\to\infty} \frac{1}{V} \int_V A(\mathbf{k}_1, \cdots, \mathbf{k}_n) e^{i[\mathbf{k}_1\mathbf{x}_1+\mathbf{k}_2\mathbf{x}_2+\cdots+\mathbf{k}_n\mathbf{x}_n]} d^{3n}\mathbf{k}.$$

Now put

$$D = \sum_{s,t=1}^{n} \nabla_s \nabla_t,$$

where

$$\nabla_s = \sum_{i=1}^{3} \frac{\partial}{\partial x_s^i} \mathbf{e}_i,$$

$\mathbf{e}_i$ independent of s, and such that $(\mathbf{e}_i, \mathbf{e}_j) = \delta_{i,j}$. Therefore

$$(-D + W)\phi(\mathbf{x}_1 \cdots \mathbf{x}_n) = \lim_{V\to\infty} \frac{1}{V} \int_V \left[\sum_{s,t=1}^{n} \mathbf{k}_s\mathbf{k}_t + W\right] A(\mathbf{k}_1, \cdots, \mathbf{k}_n) \times e^{i(\mathbf{k}_1\mathbf{x}_1+\cdots+\mathbf{k}_n\mathbf{x}_n)} d^{3n}\mathbf{k}.$$

We assume that ϕ is such that this last expression is well defined.

Let $\mathbf{k} = \sum_{s=1}^{n} \mathbf{k}_s$, $(\mathbf{k} \in R^3)$; then,

$$\operatorname{diag}([-D + W]\phi) = \lim_{V\to\infty} \frac{1}{V} \int_V (\mathbf{k}^2 + V) A(\mathbf{k}_1, \cdots, \mathbf{k}_n) e^{i\mathbf{k}\mathbf{x}} d^{3n}\mathbf{k}.$$

Let us now perform the following change of variables:

$$\sum_i \mathbf{k}_i \to \mathbf{k}, \quad \mathbf{k}_j \to \mathbf{k}_j, \quad j \neq 1.$$

Then

$$\operatorname{diag}[(-D + W)\phi](\mathbf{x}) = \lim_{V\to\infty} \frac{1}{V'} \int_{V'} [\mathbf{k}^2 + V] B(\mathbf{k}) e^{i\mathbf{k}\mathbf{x}} d^3k,$$

where

$$B(\mathbf{k}) = \lim_{V''\to\infty} \frac{1}{V''} \int_{V''} A(\mathbf{k}_1 \cdots \mathbf{k}_n) d^{3(n-1)}\mathbf{k}$$

(where V' and V'' are respectively volumes of R^3 and $R^{3(n-1)}$ such that $V'V'' = V$).

But

$$\operatorname{diag}\phi = \lim_{V'\to\infty} \frac{1}{V'} \int_{V'} B(\mathbf{k}) e^{i\mathbf{k}\mathbf{x}} d^3k,$$

hence by a trivial computation we get

$$\operatorname{diag}([-D + W]\phi)(\mathbf{x}) = (-\nabla^2 + V) \operatorname{diag}\phi(\mathbf{x}).$$

Since the eigensolutions are such that

$$(-D + W)\phi = E\phi,$$

ϕ and $(-D + W)\phi \in \mathcal{K}'$, our expression in terms of a Fourier integral is satisfied, thereby giving the first part of the theorem.

Conversely, let ψ be a function of $\mathcal{K}$ satisfying

$$(-\nabla^2 + V - E)\psi = 0.$$

We know that ψ can be written

$$\psi = \lim_{V'\to\infty} \frac{1}{V'} \int_{V'} a(\mathbf{k}) e^{i\mathbf{k}\mathbf{x}} d^3k.$$

Let us define

$$\phi(\mathbf{x}_1 \cdots \mathbf{x}_n) = \lim_{V\to\infty} \frac{1}{V'} \int_V a(\mathbf{k}_1 + \cdots + \mathbf{k}_n) e^{i(\mathbf{k}_1\mathbf{x}_1+\cdots+\mathbf{k}_n\mathbf{x}_n)} d^{3n}k.$$

We then have

$$(-D + W - E)\phi(\mathbf{x}_1 \cdots \mathbf{x}_2) = \lim_{V\to\infty} \frac{1}{V'} \int_V (\mathbf{k}^2 + W - E) a(\mathbf{k}) \times e^{i(\mathbf{k}\mathbf{x}_1+\mathbf{k}_2(\mathbf{x}_2-\mathbf{x}_1)+\cdots+\mathbf{k}_n(\mathbf{x}_n-\mathbf{x}_1))} d^{3n}\mathbf{k},$$

where, as previously, $\mathbf{k} = \sum \mathbf{k}_i$.

We make the change of variables

$$\mathbf{k} = \sum_i \mathbf{k}_i, \quad \mathbf{k}_j \to \mathbf{k}_j, \quad j \neq 1,$$

$$= (2\pi)^{3(n-1)} \lim_{V'\to\infty} \frac{1}{V'} \int_{V'} (\mathbf{k}^2 + W - E) a(\mathbf{k}) e^{i\mathbf{k}\mathbf{x}_1} d^3\mathbf{k} \times \prod_{j=2}^{n} \delta(\mathbf{x}_j - \mathbf{x}_1)$$

$$= (2\pi)^{3(n-1)} \lim_{V'\to\infty} \frac{1}{V'} \int_{V'} (-\nabla^2 + V - E) a(\mathbf{k}) e^{i\mathbf{k}\mathbf{x}_1} d^3\mathbf{k} \times \prod_{j=2}^{n} \delta(\mathbf{x}_j - \mathbf{x}_1)$$

$$= (2\pi)^{3(n-1)} (-\nabla^2 + V - E)\psi(\mathbf{x}_1) \prod_{j=2}^{n} \delta(\mathbf{x}_j - \mathbf{x}_1) = 0.$$

Corollary 1: All the eigensolutions of $-\nabla^2 + V$ are of the form

$$\psi(\mathbf{x}) = e^{i\mathbf{k}\mathbf{x}} \operatorname{diag} u_{\mathbf{k}}(\mathbf{x}),$$

where $e^{i\mathbf{k}\mathbf{x}}$ is some unitary character of R^3 and $u_{\mathbf{k}}$ is a periodic function having the same set of periods as W (or V).

Proof: If we apply Born–von Kármán boundary conditions on the solutions ϕ of $-D + W$, ϕ is of the Bloch type (in $3n$ dimensions) and the result is obvious.

Corollary 2: When W has independent periods (for each of the variables) $-D + W$ and $-\nabla^2 + V$ have the same set of eigenvalues.

[By independent periods we mean that if

$$W = \sum A_{q_1 \cdots q_n} \exp\left(i \sum_{j=1}^{n} \mathbf{q}_j \mathbf{x}_j\right),$$

then

$$\sum_{j=1}^{n} t_j^i q_j^i = 0, \quad i = 1, 2, 3,$$

where t_j^i are rational numbers, implies that $t_j^i = 0$ for all i and j.]

Proof: In view of the theorem we have only to show that diag $u \not\equiv 0$. But

$$u = \sum_{\mathbf{q}_j} B_{\mathbf{q}_1 \cdots \mathbf{q}_n} \exp\left(i \sum_j \mathbf{q}_j \mathbf{x}_j\right).$$

This implies that

$$\operatorname{diag} u = \sum_{\mathbf{q}} B_{\mathbf{q}_1 \cdots \mathbf{q}_n} \exp\left(i \sum_j \mathbf{q}_j \mathbf{x}_1\right).$$

We now have to show that if

$$\exp\left(i \sum_j \mathbf{q}_j \mathbf{x}_1\right) = \exp\left(i \sum_j \mathbf{q}'_j \mathbf{x}_1\right)$$

for any $\mathbf{x}_1$, this implies that $q_j^r = (q_j^r)'$, $\forall\, i$ and j.

But our relation implies that

$$\sum_j (q_j^i - (q_j^i)') = 0.$$

On the other hand $(q_j^i)' = t_j q_j^i$, t_j^i rational; that is, $\sum_j (t_j^i - 1) q_j^i = 0$ and, with our definition, this shows that $t_j^i = 1$ for all i and j.

Knowing the form of the solution, it would be of interest to see what remains of the band structure of such a system.

III. BAND STRUCTURE

We have seen that the general form of the solutions of

$$(-D + W)\phi = E\phi$$

are

$$\phi = \exp\left(i \sum_j \mathbf{k}_j \mathbf{x}_j\right) u_{\mathbf{k}_1 \cdots \mathbf{k}_n}(\mathbf{x}_1 \cdots \mathbf{x}_n),$$

where $\mathbf{k}_j$ and $\mathbf{x}_j \in R^3$.

Introducing this solution in the equation, we get

$$\left\{\left(\sum_j \mathbf{k}_j\right)^2 - i \sum_j \mathbf{k}_j \nabla_j - D + W - E\right\} u_{\mathbf{k}_1 \cdots \mathbf{k}_n} = 0.$$

It is clear that this operator is, by well-known arguments, Hermitian on the $3n$-dimensional unit cell.

The condition of continuity on the boundaries of the cell for the function u and its first derivatives gives a discrete set of eigenvalues E_n, for each value of $\mathbf{k} = \sum_j \mathbf{k}_j$.

E_n as a function of $\mathbf{k}$ is the analog of the energy band in the Bloch case. We have, however, here many important differences.

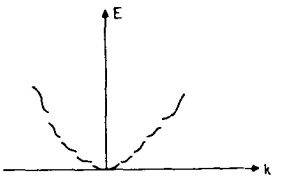

FIG. 1. Schematic bands structure of disorder systems.

First of all, $\mathbf{k}$ is now defined on the *3n-dimensional Brillouin zone* constructed on the reciprocal lattice associated with W. Since the choice of W is not unique, this shows only that we have to work here in the extended zone scheme.

Secondly, as we will show below, $E_n(\mathbf{k})$ is a function compounded of a succession of discontinuous small pieces, i.e., $E_n(\mathbf{k})$ has a very great number of gaps (see Fig. 1). (This number becomes infinite for every finite region when the periods of W are independent.)

Finally the density of states exhibit some important differences as we show in the next section.

To prove our second statement let

$$k_j^i = 2\pi p_j^i / N_j^i a_j^i,$$

where N_j^i and p_j^i are integers, and $N_j^i a_j^i$ define the Born–von Kármán condition in the $3n$-dimensional space (a_j^i basic vector of the unit cell) and also define $\tilde{k}_j^i = 2\pi / N_j^i a_j^i$ (the smallest of the K_j^i).

We assume now that we have the following conditions:

$$n_{jk}^i \tilde{k}_j^i = n_{kj}^i \tilde{k}_k^i$$

for all k, j, where n_{jk}^i and n_{kj}^i are integers.

These conditions express some *commensurability relations* between the different directions of the $3n$-dimensional lattice generated by W. In fact we have that

$$n_{jk}^i N_k^i a_k^i = n_{kj}^i N_j^i a_j^i.$$

Let $\tilde{k}_j^i = p_j^i \tilde{k}^i$, such that $\tilde{k}^i$ is as *large* as possible and let $q_j^i = 2\pi / a_j^i$ (reciprocal lattice vector).

We now have the following lemma.

Lemma 1: If $N_j^i p_j^i = \tilde{N}_j^i$ are relatively prime numbers for all the values of i and j, there exists integers s_j^i such that

$$\sum_j s_j^i q_j^i = \tilde{k}^i.$$

Proof: $\sum_j s_j^i q_j^i = \sum_j s_j^i N_j^i p_j^i \tilde{k}^i = \sum_j s_j^i \tilde{N}_j^i \tilde{k}^i$, and by the Bezout identity (Ref. 6) we can choose the s_j^i in such a way that $\sum_j s_j^i \tilde{N}_j^i = 1$.

Note that if the $\tilde{N}_j^i$ are not relatively prime, they have a highest common factor, say d^i, and the relation implies that

$$\sum_j s_j^i a_j^i = d^i k^i.$$

But since the N_j^i are very large numbers, we may assume that d^i is in general very small (except for very peculiar choices of the N_j^i which in general have no physical meaning).

As usual for a Bloch system, $E = E(k_1 + \cdots + k_n)$ is periodic with every reciprocal lattice vector $\mathbf{q}$. We now see that if we take $\mathbf{q}$ as a vector of component $s_j^i q_j^i$, E is periodic within the periods $k^i = \sum_j s_j^i q_j^i$, $i = 1, 2, 3$, that is, with very *small* periods.

Note that when the periods are independent $\mathbf{q}$ can be chosen arbitrarily near of any given quantity; that is to say that E is a constant. On the other hand E may be discontinuous for every value of the k_j^i equal to a reciprocal lattice vector. Since E cannot be a constant or nearly a constant, this implies that E is a discontinuous function.

Résumé

In summary we can say that, if the potential of a system is given by an almost periodic function, the band picture must be considered in the extended zone scheme. The energy as a function of $\mathbf{k}$ is then characterized by a great number of gaps. This number of gaps is greater if the different periods are nearly independent and becomes infinite in the limit of independent periods.

This shows that, if the variation of the energy with $\mathbf{k}$ is of the order of that of the pure crystal (i.e., finite inside of a band), the majority of the gaps must be small but do not allow us to predict the way the bands of the pure crystal are transformed when the system becomes disordered.

IV. DENSITY OF STATES

In computing the density of states, we assume the existence of a stronger relation on the commensurability conditions than that used in the last section. We require that there exists integers n_j^i such that

$$n_j^i \tilde{k}_j^i = n_k^i \tilde{k}_k^i = n_l^i \tilde{k}_l^i = \cdots .$$

The n_j^i are two by two relatively prime. More precisely we ask that in the relations of commensurability of the last section the numbers n_{jk}^i be independent of k.

These conditions mean that we have now a hypercube Ω the sides of which have length $n_j^i \tilde{k}_j^i$, and such that

$$k^i = \sum_j k_j^i, \quad i = 1, 2, 3,$$

give different values for all the set of k_j^i inside of Ω. This implies that for all different eigensolutions of $-D + W$ the diagonalization also gives different solutions of $-\nabla^2 + V$.

If we describe Ω in terms of the three-dimensional vector of components k^i, in the corresponding volume Γ, we see that the three-dimensional k-space is divided in volumes having the size of Γ, and having the same density of $\mathbf{k}$ points.

Finally, as we have seen in the last section, any $\tilde{k}_j^i$ can be described as a linear combination of components of reciprocal lattice vectors and we may, for the computation of the density of states, describe the energy in a band picture reduced to Γ.

Taking into account all these different points, we may write the density of state $g(E)$ in the form

$$g(E) = \frac{1}{(2\pi)^{3n}} \sum_{\substack{\text{branches} \\ \text{inside } \Gamma}} \frac{\partial}{\partial E} \int_0^E d^{3n}\mathbf{k}$$

and since $E = E(\mathbf{k}_1 + \cdots + \mathbf{k}_n)$ we write

$$g(E) = \frac{1}{(2\pi)^{3n}} \sum_{\substack{\text{branches} \\ \text{inside } \Gamma}} \frac{\partial}{\partial E} \int_0^E d^3k \times \int_\Gamma \prod_{i=1}^{3} \delta\left(k^i - \sum_j k_j^i\right) d^{3n}k$$

$$= \frac{1}{(2\pi)^{3n}} \sum_{\substack{\text{branches} \\ \text{inside } \Gamma}} \frac{\partial}{\partial E} \int_0^E d^3k \times \int_\Gamma \prod_{i=1}^{3} \int_{-\infty}^{+\infty} \exp\left[i\left(k^i - \sum_j k_j^i\right)x^i\right] d^{3n}k\, d^3x^i.$$

Let

$$H^i(x^i) = \int_\Gamma e^{i(k^i - \Sigma k_j^i)s^i} d^nk = 2e^{ik^i x^i}\left(\frac{\sin k_0 x^i}{x^i}\right)^n,$$

where k_0 is the side of the hypercube. Then

$$\int_{-\infty}^{+\infty} H^i(x^i)\, dx^i = 2\int_{-\infty}^{+\infty} e^{ik^i x^i}\left(\frac{\sin k_0 x^i}{x^i}\right)^n dx^i$$

$$= 4(k_0)^{n-1}\frac{n\pi}{2^n} \times \sum_{0<r<[(k^i/k_0)+n]/2} (-1)^r \left(\frac{k^i}{k_0} + n - 2r\right)^{n-1} \Big/ r!(n-r)!, \quad \text{for } 0 < k^i/k_0 < n.$$

The general expression for $g(E)$ is rather cumbersome. However, for n large enough we can use the following relation:

$$\left(\frac{\sin y}{y}\right)^n \cong 1 - \frac{n}{6}y^2 + \cdots \cong \exp -\frac{ny^2}{6}.$$

Now

$$\int_0^\infty \cos\left(\frac{k^i}{k_0}y^i\right) \exp\left(-\frac{ny^{i2}}{6}\right) dy^i = \tfrac{1}{2}\pi^{\frac{1}{2}}\left(\frac{6}{n}\right)^{\frac{1}{2}} \exp\left(-\left(\frac{k^i}{k_0}\right)^2 \frac{3}{2n}\right),$$

and $g(E)$ becomes

$$g(E) = \frac{(k_0)^{3(n-1)}}{(2\pi)^{3n} n^{\frac{3}{2}}} 8\pi^{\frac{3}{2}} \times \sum_{\text{branches}} \frac{\partial}{\partial E} \int_0^E d^3k^i \prod_{i=1}^{3} e^{-(k^i/k_0)^2(3/2n)}.$$

If we assume that E may be approached by a continuous derivable function, and put $\alpha = 3/2k_0^2 n$, we get

$$g(E) \sim \sum_{\text{branches}} \frac{\partial}{\partial E} \int_0^E d^3k e^{-\alpha k^2} = \sum_{\text{branches}} \int_E \frac{ds}{|\nabla_k E|} e^{-\alpha k^2}.$$

In the case where the periods are independent, we define

$$N(E) = g(E)/(k_0)^{3n},$$

so in this case, since $k_0 \to \infty$, we get

$$N(E) \propto \frac{\partial}{\partial E} \int_0^E d^3k.$$

This last expression is obtained by noting that

$$\lim_{k_0 \to \infty} \left(\frac{\sin k_0 x}{x}\right)^n = \pi^n [\delta(x)]^n.$$

We see that we get essentially the usual density of states but multiplied by a factor which goes to infinity with k_0.

Example: We wish to treat the case of a critical point near the edges of a band and we assume that E has a critical point E_c for $\mathbf{k} = \mathbf{k}_c$ and is such that $g(E) = 0$ for $E < E_c$. Around E_c we write

$$E = E_c + ak^2 \quad \text{and} \quad \mathbf{k} = \mathbf{k}_c + \mathbf{k}'.$$

With this we have

$$g(E) \sim \int_E k' \, d\Omega e^{-\alpha a(\mathbf{k}_c + \mathbf{k}')^2},$$

or, if $\mathbf{k}'$ is small,

$$(\mathbf{k}_c + \mathbf{k}')^2 \cong k_c^2 + 2\mathbf{k}_c\mathbf{k}'.$$

Since E_c has to be smaller than E,

$$(\mathbf{k}_c + \mathbf{k}')^2 \cong k^2 + 2k_c k',$$

and

$$g(E) \sim \left(\frac{E - E_c}{a}\right)^{\frac{1}{2}} \exp\{-\alpha k_c a[k_c - 2(|E - E_c|/a)^{\frac{1}{2}}]\}.$$

Assume that $\alpha a k_c^2$ is sufficiently large—that is, that k_c is already in the tail of the Gaussian. Then we obtain the form for $g(E)$ as shown in Fig. 2.

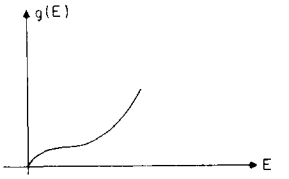

FIG. 2. Density of states near one of the edges of a band.

On the Existence of Ω (or Γ)

Assume that $V(\mathbf{x}) = \sum_n h(\mathbf{X} - \mathbf{R}_n)$, where h is a pair potential. Write now $V(\mathbf{x})$ in the following way:

$$V(\mathbf{x}) = \sum_{i=1}^{p} \sum_{n \in \Lambda_i} h(\mathbf{X} - \mathbf{R}_n),$$

where Λ_i is such that $\sum_{n \in \Lambda_i} h(\mathbf{x} - \mathbf{R}_n)$ is a periodic function (when the sum is extended to infinity). It is clear that in this case W is given by

$$W(\mathbf{x}_1 \cdots \mathbf{x}_n) = \sum_{i=1}^{p} \sum_{n \in \Lambda_i} h(\mathbf{X}_i - \mathbf{R}_n).$$

The commensurability conditions on W then imply commensurability conditions for V_i; in other words, they imply the existence of points where the different lattices corresponding to the periodic functions defined above coincide. Since physically several atoms cannot occupy the same site, this description will only be used for a *finite sample*. On the other hand, since the atoms cannot be too near each other, we see that the coincidence can happen only after several times the length of the sample has been spanned.

In mathematical terms this means that if L is the side of the sample (considered as cubic) and if

$$L = N_j^i a_j^i, \quad i = 1, 2, 3, \quad j = 1 \cdots n,$$

then for the commensurability conditions, we require that

$$n_{jk}^i N_k^i a_k^i = n_{kj}^i N_j^i a_j^i,$$

where n_{jk}^i and n_{kj}^i are integers large enough to ensure a reasonable spacing of the atoms inside the sample. These are precisely the conditions of commensurability given in Sec. 3; conditions which ensure for us the existence of k_0—that is, of the hypercube Ω.

Case of Ordered Alloys

Let us consider the case of an alloy such that if all the atoms are considered as identical we have a cubic lattice of side a.

In such case the basic vectors of our $3n$-dimensional lattice may be considered as being given by $a_j^i = p_j^i a$, where the p_j^i are integers. (This choice of the lattice is certainly not unique, but it is obviously a possible one.)

On the other hand, the generalized Born–von Kármán boundary conditions are such that

$$N_j^i p_j^i a = N_j^i a_j^i \simeq Na = L,$$

where L is the length of the sample (considered to be a cube). It follows that k_0, the side of one hypercube, is given by

$$k_0 \cong n_j^i(2\pi/Na),$$

that is, of the order of a fraction of a reciprocal lattice vector.

If we assume that the number of periods is small compared with N, $\alpha^{-1} = 2k_0^2 n/3$ will be smaller than or of the order of the smallest reciprocal lattice vector $q = (2\pi/a)$.

This shows that α^{-1} is, for the nearly free electron case, *of the order of magnitude of the band width of the pure crystal.* Further, if the band picture is not too seriously impaired in such a disordered system, we may expect a tail in the density of states for some critical value $k_c \neq 0$ corresponding to the edges of a band as shown in the previous example. The results for the density of states depend on the values of α and k_c, and on the approximation made in the calculation: first the commensurability conditions defined on a hypercube and secondly the approximation of a product of $\sin kx/x$ by a Gaussian. These different points do not allow us to consider our result as more than an attempt to understand the behavior of the density of states.

It is clearly of interest to pursue this approach further in order to elicit more information on the behavior of $g(E)$ near critical points. We consider this next.

V. THE CRITICAL POINTS OF THE DENSITY OF STATES

As in the last section, we assume that E can be described, to within a reasonable approximation, by a continuous differentiable function of the neighborhood of the critical points. We examine the density of states in the $3n$-dimensional space and so avoid dealing with any special measure in the integration.

Around a singularity, E can be expanded in Taylor series; we get

$$E = E_c + \tfrac{1}{2}\sum_{i,j,k,l}\left(\frac{\partial^2 E}{\partial k^i \partial k^j}\right)\frac{\partial k^i}{\partial k_k^i}\frac{\partial k^j}{\partial k_l^j}k_k^i k_l^j + \cdots,$$

where, as previously, $k^i = \sum_l k_l^i$, $i = 1, 2, 3$.

By a change of coordinates E can then be written

$$E = E_c + a\left(\sum_{j=0}^{p}\xi_j^2 - \sum_{j=p+1}^{n}\xi_j^2\right).$$

To obtain this last result we have used the fact that $\partial k^i/\partial k_l^i = 1$ for all l, we have diagonalized the matrix $(\partial^2 E/\partial k^i \partial k^j)_{E_c}$ and have made a homothetic transformation. (This transformation is used in the computation of Van Hove singularities.) We examine the case with $a > 0$.

First, let us perform the following change of variables:

$$\begin{aligned}
\xi_1 &= r_i \sin\theta_1 \sin\theta_2 \cdots \sin\theta_{p-1},\\
\xi_2 &= r_i \sin\theta_1 \sin\theta_2 \cdots \cos\theta_{p-1},\\
\xi_p &= r_1 \cos\theta_1,\\
\xi_{p+1} &= r_2 \sin\theta_p \cdots \sin\theta_{3n-2},\\
\xi_{3n} &= r_2 \cos\theta_p.
\end{aligned}$$

Then

$$\frac{D(\xi_1 \cdots \xi_{3n})}{D(r_1, r_2, \theta_1 \cdots \theta_{3n-2})} = (r_1)^{p-1}(r_2)^{3n-p-1}\sin\theta_1^{p-2}\cdots\sin\theta_{p-2} \times \sin\theta_p^{3n-p-2}\cdots\sin\theta_{3n-3}.$$

Put $r_1 = r\cosh\phi$, $r_2 = r\sinh\phi$, $\phi \geq 0$, from which we find

$$D(r_1, r_2, \theta_1 \cdots \theta_{3n-2})/D(r_1\phi\theta_1 \cdots \theta_{3n-2}) = r,$$

and after integration on the angles $\theta_1 \cdots \theta_{3n-2}$

$$\int_E \frac{ds}{|\nabla E|} \propto r^{3n-1}\int_{\phi\geq 0}\frac{\cosh\phi^{3n-p-1}\sinh\phi^{p-1}}{(\cosh^2\phi + \sinh^2\phi)^{\frac{1}{2}}}\,d\phi.$$

If $p = 3n$,

$$\int_E \frac{ds}{|\nabla E|} \propto r^{3n-1}$$

or

$$g(E) \propto \left(\left|\frac{E - E_c}{a}\right|\right)^{(3n-1)/2} + c, \quad \text{for } E - E_c/a > 0,$$
$$\propto c, \quad \text{for } E - E_c/a < 0.$$

The case $p = 0$ is the same apart from a change of sign for a. This shows that near the *edges* the density of states must be very steep and that the effect deduced in the previous section may be difficult to observe.

If $p \neq n$ and $\phi > 0$, since r_1 and r_2 are smaller than a certain given quantity defined by the small region chosen around E, we have for $r \to 0$

$$r_1/\cosh\phi = r_2/\sinh\phi = 0; \quad \text{that is,} \quad \phi \to \infty.$$

But for ϕ large enough, $\cosh\phi = \sinh\phi \cong e^{\phi}/2$ and

$$\int_{\phi}\frac{\cosh\phi^{3n-p-1}\sinh\phi^{p-1}}{(\cosh^2\phi + \sinh^2\phi)^{\frac{1}{2}}}\,d\phi \cong \int_{\phi}\frac{(e^{\phi})^{3n-2}}{e^{\phi}(\frac{1}{2})^{\frac{1}{2}}}(\tfrac{1}{2})^{3n-2}\,d\phi$$

such that

$$r_1, r_2 \in M = \sqrt{2}(\tfrac{1}{2})^{3n-2}\frac{(e^{\phi})^{3n-2}}{n-2} + c \cong \sqrt{2}(\tfrac{1}{2})^{3n-2}\frac{(r_1/r)^{3n-2}}{n-2} + c.$$

Therefore

$$\frac{ds}{|\nabla E|} \propto (\tfrac{1}{2})^{3n-2}\frac{r}{n-2} + c$$
$$= (\tfrac{1}{2})^{3n-2}\frac{(|E - E_c/a|)^{\frac{1}{2}}}{n-2} + c.$$

This shows that with increasing n the other singularities are washed out.

VI. LOCALIZATION OF THE ELECTRONS

In this section, we show that disordered systems, when described by almost periodic potentials, have electron states which are *more* localized than in the pure crystal. More precisely we show that, for stationary states, the wavefunction is characterized by a certain number of important peaks where the probability of presence of a localized electron is large.

We start with the equation given in Sec. I:

$$(-D + W - E)\phi = 0.$$

This is the equation we have in $3n$-dimensional space, the solution of which being such that diag $\phi = \psi$, where ψ is an eigensolution of $(-\nabla^2 + V - E)\psi = 0$.

Now write this equation in the form

$$(-D - E)\phi = -W\phi,$$

where $W\phi$ is considered as a known quantity. We proceed to solve the problem by Green's function techniques. Let G be defined by

$$-DG(\mathbf{x}, \mathbf{y}) - EG(\mathbf{x}, \mathbf{y}) = \delta(\mathbf{x}, \mathbf{y}),$$

where $\mathbf{x}, \mathbf{y} \in R^{3n}$. Put $G(\mathbf{x}, \mathbf{y}) = \int f(\mathbf{k}, \mathbf{y})e^{i\mathbf{k}x}$, where $\mathbf{k}$ is a vector of components k_j^i. The previous equation becomes

$$\int_{-\infty}^{+\infty}\left\{\sum_i\left[\left(\sum_j k_j^i\right)^2 - \epsilon\right]f(\mathbf{k}, \mathbf{y})e^{i\mathbf{k}\mathbf{x}}e^{i\mathbf{k}(\mathbf{x}-\mathbf{y})}\right\}d^{3n}\mathbf{k} = 0,$$

or, after multiplication by $e^{-i\mathbf{k}\mathbf{x}}$ and integration,

$$(\mathbf{q}^2 - E)f(\mathbf{k}, \mathbf{y}) = e^{i\mathbf{k}\mathbf{y}},$$

with

$$q^i = \sum_j k_j^i, \quad i = 1, 2, 3.$$

For $q^2 - E \neq 0$ (or $E = \epsilon + i\tau$) we have

$$G(\mathbf{x}, \mathbf{y}) = \frac{e^{-i\mathbf{k}(\mathbf{y}-\mathbf{x})}}{q^2 - E}\, d^{3n}\mathbf{k}.$$

Let us write W as a Fourier series:

$$W = \sum A_{\alpha_1^1\cdots\alpha_n^3}e^{i(\alpha_1^1x_1^1+\cdots+\alpha_n^3x_n^3)}$$
$$= \sum a_{\alpha_1^1}\ldots a_{\alpha_n^3}e^{i(\alpha_1^1x_1^1+\cdots+\alpha_n^3x_n^3)},$$

where the $a_{\alpha_j^i}$ are chosen in a convenient manner such that

$$\prod a_{\alpha_j^i} = A_{\alpha_1^1\cdots\alpha_n^3}.$$

Then

$$W = \prod_{i=1}^{s}\prod_{j=1}^{n}\left(\sum_{\alpha_j^i} a_{\alpha_j^i}e^{i\alpha_j^ix_j^i}\right)$$

and if we put

$$W_j^i(x_j^i) = \sum_{\alpha_j^i} a_{\alpha_j^i}e^{i\alpha_j^ix_j^i},$$

then W_j^i is a periodic function. We now divide W_j^i into a sum of functions

$$W_j^i(x_j^i) = \sum_{R_j^i} V_j^i(x_j^i - R_j^i).$$

(The R_j^i are the lattice sites.)

The V_j^i are certainly not uniquely defined but it is clear that such a decomposition always exists.

With this result we have for ϕ the following expression:

$$\phi(\mathbf{y}) = \int\cdots\int\frac{e^{i\mathbf{k}(\mathbf{y}-\mathbf{x})}}{q^2 - E}$$
$$\times\prod_{i=1}^{3}\prod_{j=1}^{n}\left[\sum_{R_j^i} V_j^i(x_j^i - R_j^i)\right]\phi(\mathbf{x})d^{3n}kd^{3n}x,$$

or, writing $x_j^i - R_j^i = z_j^i$,

$$\phi(\mathbf{y}) = \int\cdots\int\frac{1}{q^2 - E}$$
$$\times\prod_{i=1}^{3}\prod_{j=1}^{n}\left[\sum_{R_j^i} e^{ik_j^i(y_j^i - z_j^i - R_j^i)}V_j^i(z_j^i)\right]$$
$$\times\,\phi(\mathbf{z} + \mathbf{R})d^{3n}kd^{3n}z.$$

By the Corollary 1 of Sec. II we have

$$\phi(\mathbf{x}) = e^{i\mathbf{k}'\mathbf{x}}u_{\mathbf{k}'}(\mathbf{x})$$

and

$$\phi(y) = \int\cdots\int\frac{1}{q^2 - E}\prod_{i=1}^{3}\prod_{j=1}^{n}\left[\sum_{R_j^i} e^{ik_j^i(y_j^i - z_j^i - R_j^i)}V_j^i(z_j^i)\right]$$
$$\times\, e^{i\mathbf{k}'(\mathbf{z}+\mathbf{R})}u_{\mathbf{k}'}(\mathbf{z})d^{3n}kd^{3n}z.$$

Expressing $u_{k'}(\mathbf{x})$ in a manner similar to W, we obtain

$$\phi(\mathbf{y}) = \int\cdots\int\frac{1}{q^2 - E}\prod_{i=1}^{3}\prod_{j=1}^{n}\left[\sum_{R_j} e^{i(k_j'^i - k_j^i)R_j^i}\right.$$
$$\left.\times\, e^{i[(k_j'^i - k_j^i)z_j^i + k_j^iy_j^i]}V_j^iu_j^i(z_j^i)\right]d^{3n}kd^{3n}z.$$

However,

$$\sum_{R_j^i} e^{i(k_j^{i\prime} - k_j^i)R_j^i} = N_j^i\sum_{p_j^i}\delta(k_j^ip_j^i - k_j'^i),$$

where the p_j^i are reciprocal lattice vector components. So that writing

$$N = \prod_{ij} N_j^i$$

and

$$q'^i = \sum_{j=1}^{n}(k_j'^i + p_j^i),$$

we get

$$\phi(y) = N \sum_p \int \cdots \int \frac{1}{q'^2 - E} \times e^{i(p_j^i z_j^i + (k_j^{i\prime} + p_j^i) y_j^i)} V_j^i u_j^i(z_j^i) d^{3n} z.$$

Let

$$u_j^i V_j^i(z_j^i) = \int a_j^i(s_j^i) e^{-i s_j^i z_j^i} \, ds_j^i.$$

Then

$$\phi(\mathbf{y}) = N \sum_p \frac{1}{q'^2 - E} \prod_{i=1}^{3} \prod_{j=1}^{n} a_j^i(p_j^i) e^{i \sum_{ij} (k_j^{i\prime} + p_j^i) y_j^i},$$

so that after diagonalization we get

$$\operatorname{diag} \phi(\mathbf{x}) \equiv \psi(\mathbf{x}) = N \sum_p \frac{1}{q'^2 - E} \prod_{i=1}^{3} \prod_{j=1}^{n} a_j^i(p_j^i) e^{i \mathbf{q}' \mathbf{x}}.$$

In principle we can now put ϕ in the equation $(-D + W - E)\phi = 0$ and determine by a determinantal equation both $\prod_{ij} a_j^i$ and the energy. Since in general this is very difficult to perform, we first assume some special form for the function $u_j^i V_j^i$, functions which can be regarded as rapidly decreasing outside the unit cell centered at the origin.

(1) We assume that

$$u_j^i V_j^i(x_j^i) = \frac{\lambda_j^i}{2}, \quad -\alpha_j^i < x_j^i < \alpha_j^i,$$
$$= 0 \text{ elsewhere},$$

which is a kind of generalized muffin-tin potential: i.e.,

$$a_j^i(s_j^i) = \frac{\lambda_j^i}{2} \int_{-\alpha_j^i}^{\alpha_j^i} e^{i s_j^i z_j^i} \, dz_j^i = \lambda_j^i \frac{\sin s_j^i \alpha_j^i}{s_j^i}.$$

But

$$\prod_{i=1}^{3} \prod_{j=1}^{n} \lambda_j^i \frac{\sin s_j^i \alpha_j^i}{s_j^i} \cong \prod_{ij} (\lambda_j^i \alpha_j^i) \exp\left[-\sum_{lk} (s_k^l \alpha_k^l)^2 / 6 \right]$$

for n large enough. Then

$$\phi(y) = N \sum_p \frac{\exp\left(-\sum_{ij} (p_j^i \alpha_j^i)^2 / 6 \right)}{q'^2 - E} e^{i(\mathbf{k}' - \mathbf{p})\mathbf{y}},$$

where $\lambda = \prod_i \lambda_j^i \alpha_j^i$.

(2) Take

$$u_j^i V_j^i(x_j^i) = -\lambda_j^i \frac{1}{(x_j^i)^2 + (b_j^i)^2}, \quad b_j^i > 0.$$

Since

$$\int_{-\infty}^{+\infty} \frac{1}{x^2 + b^2} e^{isx} \, dx = \pi b^{-1} e^{-b|s|},$$

we get

$$\prod_{ij} a_j^i(s_j^i) = \lambda e^{-\sum_{ij} b_j^i |p_j^i|}$$

and

$$\phi(\mathbf{y}) = N\lambda \sum_p \frac{e^{-\sum_{ij} b_j^i |p_j^i|} e^{i(\mathbf{k}' - \mathbf{p})\mathbf{y}}}{q'^2 - E}.$$

(3) Take

$$u_j^i V_j^i(x_j^i) = -\lambda_j^i \frac{\alpha_j^i + x_j^i}{(p_j^i)^2 + (\alpha_j^i + x_j^i)^2} - \frac{\alpha_j^i - x_j^i}{(p_j^i)^2 + (\alpha_j^i - x_j^i)^2}, \quad \text{with} \quad p_j^i > 0.$$

We obtain

$$\prod_{ij} a_j^i(s_j^i) = \prod_{ij} - \lambda_j^i [\pi e^{-\beta_j^i |s_j^i|} \sin(\alpha_j^i |s_j^i|)]$$

and

$$\phi(\mathbf{y}) = cs + \sum_p \prod_{ij} \sin(\alpha_j^i |p_j^i|) \exp \frac{e^{-\sum_{ij} \beta_j^i p_j^i} e^{i(\mathbf{k} - \mathbf{p})\mathbf{y}}}{q'^2 - E}$$

In the general case we note that since $\|u_j^i V_j^i\| < \infty$, the series of the $|a_j^i(p_j^i)|$ must converge. Then $\prod |a_j^i(p_j^i)|$ must converge very rapidly with increasing n.

The main differences between the general expressions for the eigensolutions of $-\nabla^2 + V$ for a periodic potential or for an almost periodic potential can now be investigated both in the examples and in the general case:

(a) For the case $n = 1$ (periodic), $\psi(\mathbf{x}) \equiv \operatorname{diag} \phi(\mathbf{x})$ is a Bloch type function, where the periodic part is expanded in Fourier series."

For $n > 1$ ψ is a "generalized Bloch" type solution where the almost periodic part is also developed in terms of a "generalized Fourier series."

(b) As we have seen in Lemma 1 of Sec. III, if $q_j^i = 2\pi / a_j^i$ (reciprocal lattice vector), then there exists an integer s_j^i such that $\sum_{s_j^i} s_j^i q_j^i = 2\pi / N_j^i a_j^i$. This shows that q'^i and consequently $\mathbf{q}'^2$ can approach any given quantity to within $2\pi / N_j^i a_j^i$—that is, physically arbitrarily near. But this is true only in the case $n > 1$.

It follows also in this latter case that the p_j^i can be chosen, *for positive energy*, in such a way that $q'^2 - E \cong 0$.

We can now distinguish several cases. Let δ be a small set of values of $\mathbf{q}'$ around the value $\mathbf{q}_0'$ defined be $q_0'^2 = E$. Then we may have the following.

(1) $a_j^i(p_j^i)$ is of the same order of magnitude for the p_j^i for which $\mathbf{q}' \in \delta$, and $\psi(\mathbf{x})$ is dominated by

$$\sum_{\substack{p \\ q' \in \delta}} \frac{1}{q'^2 - E} \exp\left[i \sum_{i=1}^{3} \left(\sum_{j=1}^{n} (k_j^i - p_j^i) x^i \right) \right].$$

That is, we have a superposition of plane waves with similar amplitudes and small changes in phase. Then $|\psi(\mathbf{x})|$ will be very much *localized in space*; more precisely $|\psi(\mathbf{x})|$ will be characterized by a series of peaks.

We may observe that, since $\prod a_j^i(p_j^i)$ converges rapidly, for high energy δ will rapidly be reduced to a

few points or to a single point, leading to a single plane wave, as expected. *So the effect of localization is more to be expected at low energy.*

This situation is illustrated by examples (1) and (2), if we assume that respectively $\alpha_j^i p_j^i$ and $b_j^i p_j^i$ are of the same order at magnitude for $j = 1 - n$ (p fixed).

(2) If $\prod_{ij} a_j^i(p_j^i) = 0$ for the p_j^i corresponding to elements of δ, or if $E < 0$, then the behavior is essentially defined by the function $\prod_{ij} a_j^i(p_j^i)$ in other regions.

This will lead to smooth peaks by the cancellation of certain phases. But the localization will be much less concentrated than in the previous case. It is worth noticing, however, that *if* $E > 0$ this situation will not happen very often. In fact, if $u_j^i V_j^i$ is an even distribution and if we assume, without loss of generality, that $u_j^i V_j^i$ is of compact carrier (This means essentially that the functions $u_j^i V_j^i$ are zero outside a given closed set. Note that the freedom we have in the division of the u_j^i always allows us to take u_j^i of compact carriers.), then by the theorem of Paley Wiener[7] the Fourier transform of $u_j^i V_j^i$ is an entire analytic function of exponential type. This implies that the zeros of the function $\prod a_j^i$ are at least enumerable and so δ has always some points giving an important contribution to $\psi(\mathbf{x})$.

This situation is illustrated by the last example where the function sinus may be zero for some of the p_j^i.

APPENDIX: SOME PROPERTIES OF ALMOST PERIODIC FUNCTIONS (APF)

Let G be a commutative locally compact group and $\hat{G}$ the set of all characters of G. If we define the product of two characters, χ and χ', by $\chi\chi'(s) = \chi(s)\chi'(s)$, $\hat{G}$ becomes a group which, provided with the topology of compact convergence, is called the *dual group* of G. Let now G_d be the dual of G provided with the discrete topology, and G' be the dual of G_d. One can then show (3) that G' is compact. Moreover G can be canonically injected into G'. G' is called the compact group associated with G.

Proposition 1: Let f be a continuous bounded function on G'. The following conditions are equivalent.

(1) f is the restriction to G of a continuous function on G'.

(2) f is a uniform limit of linear combination of characters.

(3) The set of all the translated functions f_s of f is relatively compact in $C^\infty(G)$ (for the topology of the uniform convergence). For proof see Ref. 2, p. 91.

Definition: A continuous bounded function which verifies the conditions given in the proposition (1) is called almost periodic. (This definition can be generalized to general locally compact groups. See Ref. 3, p. 298.)

We now want to take for G the group of the translations of R^3. For this group we have the following proposition.

Proposition 2: A continuous bounded function f on R^3 will be almost periodic if and only if for every $\epsilon > 0$ there exists $T > 0$ ($T \in R$) such that every ball of diameter T contains the extremity of a vector $\boldsymbol{\tau}$ satisfying

$$\|f - f_\tau\|_\infty \leq \epsilon.$$

Proof: We repeat here (with the trivial necessary changes) the proof given in the case of R in Ref. 2, p. 95. Since f is periodic, the set of f_s is relatively compact then precompact—that is to say, that for every $\epsilon > 0$ there exists $\mathbf{a}_1 \cdots \mathbf{a}_n \in R^3$, such that for all $\mathbf{s} \in R^3$ one can find an $i = 1 \cdots n$ for which

$$\|f_s - f_{a_i}\| \leq \epsilon.$$

But this is equivalent to $\|f_{s-a_i} - f\| \leq \epsilon$. If we now put $T = 2 \sup |a_i|$, then for any point $\mathbf{s}$ the ball of diameter T contains vectors which satisfy the condition.

Conversely if the conditions are satisfied,

$$\sup |f(\mathbf{x} + \mathbf{y}) - f(\mathbf{x})| \leq \sup |f(\mathbf{x} + \mathbf{y}) - f(\mathbf{a})| + \sup |f(\mathbf{x}) - f(\mathbf{a})| \leq 2\epsilon.$$

So f is uniformly convergent.

Now for all $t \in R^3$ there exists $\mathbf{s} \in$ the ball A of diameter T centered on the origin such that

$$\|f - f_{t-s}\| \leq \epsilon \quad \text{or} \quad \|f_s - f_t\| \leq \epsilon.$$

But the set f_s, $\mathbf{s} \in A$, is compact (for a continuous mapping of a metric space into a metric space, a compact set goes onto a compact set), so there exists $b_1 \cdots b_p \in A$ such that for all $\mathbf{s} \in A$ we can find an $i = 1 \cdots p$ such that $\|f_s - f_{b_i}\| < \epsilon$; then

$$\|f_t - f_{b_i}\| \leq \|f_t - f_s\| + \|f_s - f_{b_i}\| \leq 2\epsilon$$

and the set of the translation is relatively compact.

Fourier Transform of an APF

We define the Fourier transform of an APF, f, as the Fourier transform of the function f' defined as the restriction of f to the compact group G' associated with G. Since G' is compact (Ref. 2, p. 90), the Fourier transform of f tends to zero at infinity. This implies that this function is different from zero only on an

enumerable set. This set noted Λ is called *the set of frequencies of* f.

In the case of $G = R^3$, the characters are the usual exponentials of modulus 1. We see that by the definition of f, f is expressed as a series of such exponentials,

$$f(\mathbf{x}) = \sum_{\lambda} A_\lambda e^{i\lambda\mathbf{x}}.$$

In the general case we get

$$f(g) = \sum_{\lambda\in\Lambda} A_\lambda x^\lambda(g), \quad x^\lambda \in \hat{G}.$$

Connection of APF on R^3 with Periodic Functions of Several Variables

Definition 1: A set α_i, $i = 1, 2, \cdots$, of real numbers will be called independent if for any n the only rational values of $r_1 r_2 \cdots r_n$ satisfying the equation

$$r_1\alpha_1 + \cdots + r_n\alpha_n = 0 \quad \text{are} \quad r_1 = r_2 \cdots = r_n = 0.$$

Let $f(\mathbf{x}) = \sum A_\lambda e^{i\lambda\mathbf{x}}$.

Definition: A finite or enumerably infinite set α_i of linearly independent numbers is called a basis of the function $f(x)$ if every λ^j (component of $\boldsymbol{\lambda}$) can be represented as a finite linear form of α's with rational coefficients:

$$\lambda^j = r_1^{(j)}\alpha_1 + \cdots + r_n^{(j)}\alpha_n;$$

n depends on j and λ.

Evidently every APF has a basis, and there may be many different bases of the same function, but for a fixed basis the representation is unique.

We will assume from now on that the set α_i of "fundamental" frequencies is always *finite* or that the basis we use is finite. In fact this restriction can be dropped by using some limiting processes [see Ref. 1, p. 35].

With the use of the preceding definition the Fourier expansion of any APF take the form

$$(+) \quad f(\mathbf{x}) = \sum_{r_j^i} A_{r_1^1\cdots r_n^1\cdots r_n^3} e^{i(\sum_{jk} r_j^k \alpha_j x^k)} = \sum_{\mathbf{r}} A_{\mathbf{r}_1\cdots\mathbf{r}_n} e^{i\sum_j \alpha_j(\mathbf{r}_j\mathbf{x})}.$$

Taking into account Definition 1 of Sec. II, we have the following.

Proposition: For any APF f, there exists a function F at several variables, F *periodic in each variable*, such that

$$f = \text{diag } F.$$

Proof: Let f be given by (+) and define F by

$$F(\mathbf{x}_1, \cdots, \mathbf{x}_n) = \sum_{\mathbf{r}} A_{\mathbf{r}_1\cdots\mathbf{r}_n} e^{i(\mathbf{r}_1\alpha_1\mathbf{x}_1 + \mathbf{r}_2\alpha_2\mathbf{x}_2 + \cdots + \mathbf{r}_n\alpha_n\mathbf{x}_n)}.$$

This function is well defined since $\sum_r |A_{\mathbf{r}_1\cdots\mathbf{r}_n}|^2 < \infty$ and diag $F = f$.

The Pre-Hilbert Space of APF

We can show (see Ref. 2, p. 94) that almost periodic functions generate a pre-Hilbert space—that is, a set of functions the closure of which is a Hilbert space.

In the case of R^3, the scalar product is given in this space by

$$(f \mid g) = \lim_{V\to\infty} \frac{1}{V} \iiint_V \overline{f(x)} g(\mathbf{x}) d^3\mathbf{x}.$$

On the other hand the Fourier transform $\tilde{f}$ of f is given by

$$\tilde{f}(\mathbf{k}) = \lim_{V\to\infty} \frac{1}{V} \iiint_V f(\mathbf{x}) e^{i\mathbf{k}\cdot\mathbf{x}} d^3x.$$

CONCLUSION

Our model of disordered systems described by an APF potential allows us to obtain some general features on the behavior of such systems. It is easy to see that the same techniques can be used to include an electromagnetic field. In this case it seems that we would be able to describe the transport properties of electrons moving in potentials represented by APF's.

Finally it is possible to use the $3n$-dimensional description we have introduced to describe the x-ray scattering for the systems under consideration.

ACKNOWLEDGMENTS

The author should like to express his sincere thanks to Prof. N. W. Ashcroft for very helpful comments and discussions, and for a critical reading of the manuscript. He wishes to thank Dr. J. Krumhansl for his interest in this work and for acquainting him with literature on similar subjects. Finally he is grateful to Dr. K. Sturm for some valuable comments.

* Supported in part by Swiss National Scientific Research Fund and in part by National Science Foundation.

[1] A. S. Besicovitch, *Almost Periodic Functions* (Cambridge, London, 1932).
[2] A. Guirchardet, *Analyse harmonique commutative* (Dunod, Paris, 1968).
[3] J. Dixmier, Les C^*-algèbres et leurs représentations (Gauthier-Villards, Paris, 1964).
[4] B. I. Halperin and M. Lax, Phys. Rev. **148**, 724 (1966).
[5] J. Zittartz and J. S. Langer, Phys. Rev. **148**, 741 (1966).
[6] See, for example, B. L. van der Waerden, *Modern Algebra* (Ungar, New York, 1949).
[7] L. Schwartz, *Théorie des distributions* (Hermann, Paris, 1966), p. 272.

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 12, NUMBER 3 MARCH 1971

Equations of Motion in Nonequilibrium Statistical Mechanics. III. Open Systems

BALDWIN ROBERTSON AND WILLIAM C. MITCHELL*
National Bureau of Standards, Washington, D.C. 20234

(Received 10 August 1970)

A simple hypothesis on the effect of the interaction between a system and its surroundings is used to generalize nonequilibrium statistical mechanics to apply to open systems. Thermal driving of a system by its surroundings is defined in statistical mechanics by analogy with the first law of thermodynamics, which describes exchange of heat between the system and an external source. The assumption that an isolated system is thermally driven is used to derive a Liouville equation with an additional term that is linear in the external source strength. The generalized Liouville equation is used to derive closed equations of motion that are the same as for an isolated system except for an additional term, which is just the source strength. This formalism is attractive because the source strength, which is assumed known, appears in the equations linearly just as in classical thermodynamics or hydrodynamics. A microscopic expression for the source strength is obtained by comparing the thermal driving formalism with an exact dynamical analysis of the system interacting with its surroundings.

1. INTRODUCTION

This paper presents a simple hypothesis on the effect of the interaction between a system and its surroundings. This hypothesis is used to derive a generalization of the Liouville equation for open systems, which is then used to derive closed equations of motion for the physical variables of the system. The equations of motion are also derived exactly by considering the system and its surroundings together as an isolated system, and the results are compared.

In thermodynamics, changes in the macroscopic state of a system are of two kinds. A system can be dynamically driven by doing work on it, or thermally driven by letting heat enter or leave it. Dynamical driving has an obvious microscopic definition and hence fits naturally into statistical mechanics. Thermal driving, on the other hand, has no obvious microscopic definition and hence must be formulated as an additional hypothesis in statistical mechanics.

The thermal driving hypothesis that we make is simply that the physically observable macroscopic effects of the interaction between the system and its surroundings are sufficient, along with the Hamiltonian of the system, to determine the time development of the density operator of the system. Specifically, thermal driving is described in our formalism by just the macroscopic external source strength, the macroscopic state of the system of interest, and the operators whose expectations express this state. All the details of the interaction with the surroundings are contained in the source strength, which is a macroscopic parameter that is directly observed. This definition of thermal driving is formulated by analogy with the first law of thermodynamics, with the dQ term becoming the source strength. In this manner, the parallelism between dynamical and thermal driving in thermodynamics is incorporated into nonequilibrium statistical mechanics.

The concept of thermal driving is useful when the macroscopic state of the system does not depend upon the details of the interaction Hamiltonian, but does depend upon the external source strength. This is a common situation. For example, the state of the water in a tea kettle on a stove depends upon how fast heat is added and how fast the water evaporates, but not upon the kind of stove used or where the steam goes after it leaves the surface of the water.

The assumption that an isolated system is thermally driven leads immediately to a generalization of the Liouville equation for open systems. This equation differs from the Liouville equation for isolated systems only by an additional term that is linear in the external source strength and otherwise depends only upon the macroscopic state of the system of interest and upon the operators whose expectations express this state. The generalized Liouville equation is used to derive closed equations of motion for the expectations of these operators. These equations of motion are exactly the same as those for an isolated system except for an additional term, which is just the external source strength.

The equations of motion derived previously[1,2] for an isolated system are exact partial integro-differential equations whose kernels are time-correlation functions. The expectations of the operators corresponding to the quantum-mechanical variables describing the macroscopic state of the system are related by another set of equations to the thermodynamic conjugate variables such as temperature, chemical potential, etc. These equations and the equations of motion, taken together, are closed in the sense that the expectations and their thermodynamic conjugates are the only

unknowns, although some of the dependence in the equations on the conjugates is expressed only implicitly. The equations of motion are exact memory-retaining nonlocal generalizations of the equations of nonequilibrium thermodynamics. Hence the equations can describe thermal relaxation and transport in an isolated system. In the Markovian local limit the equations reduce to the equations of nonequilibrium thermodynamics, where the transport coefficients are time and space integrals of the kernels.

An introductory discussion of this formalism has been given in the first section of Ref. 2. The formalism has been applied to the derivation of the equations of motion of nuclear magnetism[3] and to the derivation of generalized hydrodynamic equations from quantum statistical mechanics.[4] This last paper is self-contained and hence also serves as an introduction.

In the present paper on open systems, the main text, but not all the footnotes, can be read without a previous reading of these references.

The remainder of this paper in outline is as follows. In the rest of this section several other papers on open systems are discussed. In Sec. 2 the definition of thermal driving is given[5] and is used to derive the generalization of the Liouville equation for open systems. This equation is used to find the corresponding generalization of the equations of motion. In Sec. 3 transport into a system is analyzed purely dynamically by deriving the exact equations of motion for a composite system consisting of the system of interest and its surroundings. This formalism is compared with the thermal driving formalism of Sec. 2, and a microscopic expression for the source strength is obtained. This expression is used to discuss limitations in the usefulness of the thermal driving formalism for open systems.

Open systems have also been discussed by other authors, whose work we now relate to ours.

In order to obtain expressions for the transport coefficients, McLennan[6] considers a nonequilibrium process that results from the nonconservative forces exerted on the system of interest by a number of external reservoirs. He obtains two different expressions involving the effects of the reservoirs, and then eliminates between them the reservoir parts. He thus obtains an expression for a nonequilibrium statistical density that depends only upon the conjugate variables of the system alone, with no reference to the reservoirs. The reservoirs were not even needed in the derivation since, as Zubarev[7] pointed out, this nonequilibrium statistical density is easily shown to satisfy the Liouville equation for the isolated system alone. Thus, McLennan was interested in using the effect of the reservoirs on the system only to obtain a description of a nonequilibrium process within the system.

In order to obtain a description of nonequilibrium processes, Bergmann and Lebowitz[8] add a term to the Liouville equation to represent the effect of the surroundings. The additional term in the Liouville equation that we derive cannot be expressed in the simple form that they assume.

Bergmann and Lebowitz[8] also suggest that a system can approach equilibrium only by interacting with its surroundings. Indeed, it is true that once an isolated system is disturbed dynamically from equilibrium, its statistical density can never approach a canonical distribution. However, macroscopic systems can be described by a relatively small number of variables, and an isolated system can approach a state in which these variables take values computed using a canonical distribution. This approach to equilibrium in an isolated system is described by the formalism of Refs. 1 and 2, and the present paper is merely an extension of that formalism to the more general case of nonconservative interaction with the surroundings.

Emch and Sewell[9] consider a system interacting dynamically with a number of infinite reservoirs each in internal equilibrium and obtain an equation of motion for the statistical density operator of the system alone. The effect of the surroundings in their generalized Liouville equation is given by a memory integral term, whose kernel is a system operator that depends on the reservoir temperatures and on the interaction Hamiltonian. In the present paper the hypothesis of thermal driving is used to express the physically relevant part of this interaction in terms of a macroscopic external source strength, which does not appear in a memory integral.

None of these authors used their generalized Liouville equations to obtain closed equations of motion for the system's physical variables of interest. Furthermore, in their work, the external source strength does not appear explicitly, and hence the correspondence with thermodynamics is not apparent.

2. STATISTICAL MECHANICS OF THERMALLY DRIVEN SYSTEMS

Let $F_1(\mathbf{r}), F_2(\mathbf{r}), \cdots, F_m(\mathbf{r})$ denote the linearly independent quantum mechanical (i.e., Hermitian, linear, and possibly noncommuting) operators whose expectations correspond to the macroscopic variables that are observed or controlled in the experiment considered. For example, these operators may be the energy density and particle density, which can depend upon

position **r**. In addition to these variables there are their thermodynamic conjugates; all together they are assumed to form a thermodynamically complete set of variables. However, for the present we need to consider only the operators and their expectations. In order to simplify notation, we will suppress the subscript and the **r** and write just F to represent these operators in this paper.

When an isolated system is dynamically driven by its surroundings, its Hamiltonian $\mathcal{H}(t)$ has a time dependence due to the time dependence of the external force fields acting on the system. This Hamiltonian appears in the Liouville equation for the statistical density operator $\rho(t)$. The expectation of F, then, is calculated in principle by solving this equation for $\rho(t)$, multiplying $\rho(t)$ by F, and taking the trace. As a result this expectation will have a time dependence, which we say is due to the dynamics.

When the system is thermally driven, however, its variables that are observed or controlled are changed not only dynamically, but also thermally by addition to these variables from the surroundings, while all of its variables other than those observed or controlled are changed only dynamically and not changed directly by the surroundings. Let G represent all possible operators of the system other than those corresponding to variables that are observed or controlled. Also let $f(t)$ be the rate of change that is directly due to additions from the surroundings. Like F, the source strength $f(t)$ also has a subscript and a dependence on **r** that has been suppressed. With this notation the definition of thermal driving[5] is

$$\frac{\partial \operatorname{Tr}[F\rho(t)]}{\partial t} - \operatorname{Tr}[\dot{F}\rho(t)] = f(t), \tag{1}$$

$$\frac{\partial \operatorname{Tr}[G\rho(t)]}{\partial t} - \operatorname{Tr}[\dot{G}\rho(t)] = 0. \tag{2}$$

These equations state that the total rate of change of the expectations of F and G minus the rate of change due to the dynamics equals the rate of change due to additions from the surroundings. The first equation has the same form as the first law of thermodynamics, $dE + dW = dQ$. Because the second equation has a zero on the right, it is necessary for F to include all the operators corresponding to the system variables that have additions from the surroundings. All the variables corresponding to the operators represented by G are assumed not to have additions from the surroundings.

The new feature in these equations is the presence of the $f(t)$ terms, which implies a new equation of motion for $\rho(t)$. This generalizes nonequilibrium statistical mechanics to correspond more closely with thermodynamics, which can describe, for example, heat or particles added to a system as well as work done on a system by its surroundings.

In order to use these equations, it is necessary to give the operators represented by G a more precise definition. Since G corresponds to all those variables that are not observed or controlled, it must be orthogonal to F, which corresponds to those that are. This orthogonality can be expressed by

$$G = [1 - \mathcal{P}(t)]A, \tag{3}$$

where A is an arbitrary quantum mechanical operator of the system and $\mathcal{P}(t)$ is an operator that gives the projection of A along F. This projection operator is defined by[10]

$$\mathcal{P}(t)A \equiv \langle A\bar{F}\rangle_t \cdot \langle F\bar{F}\rangle_t^{-1} \cdot F. \tag{4}$$

The angular brackets are defined by

$$\langle A\rangle_t \equiv \operatorname{Tr}[A\sigma(t)], \tag{5}$$

where $\sigma(t)$ is a statistical density operator that depends upon $\rho(t)$. We do not need to know that dependence right now; all we need to know now is that $\sigma(t)$ is some given function of $\rho(t)$. Also the bar[11] over F in Eq. (4) can be ignored for now.

The dot over the operators F and G in Eqs. (1) and (2) is defined by

$$\dot{A} \equiv \frac{\partial A}{\partial t} + \frac{i}{\hbar}[\mathcal{H}(t), A], \tag{6}$$

where A is either F or G. These are Schrödinger operators, and so the first term on the right of Eq. (6) will be zero for F, but not for G because of the time dependence given by Eqs. (3)–(5).

In general, F represents several spatially dependent operators and so has a subscript and a dependence on **r**, which we have suppressed. Hence $\langle F\bar{F}\rangle_t$ is a matrix kernel, which depends upon two space coordinates, and the $\langle F\bar{F}\rangle_t^{-1}$ in Eq. (4) is its inverse.[12] The dot in Eq. (4) has the following special meaning. Say we have some (possibly time-dependent) functions $\lambda_1(\mathbf{r}, t), \cdots, \lambda_m(\mathbf{r}, t)$, which we represent by just $\lambda(t)$. Then a dot between $\lambda(t)$ and F represents the sum and integral:

$$\lambda(t) \cdot F \equiv \sum_{n=1}^{m} \int d^3r \lambda_n(\mathbf{r}, t) F_n(\mathbf{r}).$$

In Eq. (4) there are two such sums and integrals over the subscripts and space dependences of the operators F.

Equations (1)–(6) give our definition of thermal driving, where the A in Eq. (3) is to range over all possible system operators.

These equations are sufficient to determine the equations of motion for $\rho(t)$. By inserting Eq. (6) into Eq. (2), cancelling a term, inserting Eqs. (3) and (4) into the result, and using Eqs. (1), (6), and (5), we get

$$\mathrm{Tr}\left(A\frac{\partial\rho}{\partial t}+\mathrm{Tr}\left(A\,\frac{i}{\hbar}\,[\mathcal{H}(t),\rho(t)]\right)\right.$$
$$=\mathrm{Tr}\,[A\bar{F}\sigma(t)]\cdot\langle F\bar{F}\rangle_t^{-1}\cdot f(t).$$

Since this must be true for arbitrary A, we get[13]

$$\frac{\partial\rho(t)}{\partial t}+\frac{i}{\hbar}\,[\mathcal{H}(t),\rho(t)]=\bar{F}\sigma(t)\cdot\langle F\bar{F}\rangle_t^{-1}\cdot f(t). \quad (7)$$

This is the equation of motion for $\rho(t)$; it is closed since $\sigma(t)$ is a given function of $\rho(t)$ and since $f(t)$ is assumed given. This generalization of the Liouville equation is attractive because it is Markovian and because the source strength $f(t)$ appears in it explicitly.

Equation (7) can be used to obtain equations of motion for the expectations of F in a way similar to that used in Ref. 1. There the equations of motion were derived using a statistical density operator that is canonical in the operators F, with coefficients $\lambda(t)$ chosen to give the correct expectation for F. For thermally driven systems, the equations of motion will be closed only if we choose the $\sigma(t)$ in the above definition of thermal driving to be just this generalized canonical density operator. This choice now completes our definition of thermal driving.[14] This definition and the resulting generalized Liouville equation thus depend not only on which thermodynamic coordinates of the system are directly changed by the surroundings, but also on what additional operators must be included in F in order to obtain a complete thermodynamic description of the isolated system. Once F is chosen, the effect of the nonconservative interaction with the surroundings is given in our formalism in terms of just the external source strength $f(t)$, the system's conjugate variables $\lambda(t)$, and the operators F. The details of the interaction are irrelevant.

With this definition of $\sigma(t)$, the generalized Liouville equation (7) can now be used as before[15] to derive the equation of motion for $\langle F\rangle_t$, and the result is

$$\frac{\partial\langle F\rangle_t}{\partial t}=\langle\dot{F}\rangle_t+f(t)+\int_0^t dt'K(t,t')\cdot\lambda(t'). \quad (8)$$

Here $\langle\dot{F}\rangle_t$ is the reversible change in $\langle F\rangle_t$, the kernel $K(t,t')$ is the autocorrelation of $\dot{F}$, and $\lambda(t)$ is the thermodynamic conjugate to $\langle F\rangle_t$. All of these are exactly the same functionals of $\langle F\rangle$ as before[16] so that the equation of motion is still closed. The only difference is that the external source strength $f(t)$, which is presumed known, is simply added to the equation. This is exactly the way an external source strength appears in the classical hydrodynamic equations.

This simple result occurs only because the $\sigma(t)$ used in the definition of thermal driving is the same as the generalized canonical density used in deriving the equation of motion. Otherwise the equation would not be closed.

Thus our definition (1)–(6) of thermal driving is attractive not only because it itself is the statistical mechanical analog of thermal driving in thermodynamics, but also because it easily gives a simple generalization of the Liouville equation (7), which leads directly to the closed equation of motion (8). In all these equations the source strength $f(t)$ appears explicitly and not in a memory integral, and in Eq. (8) it appears exactly as in the classical hydrodynamic equations.

3. COMPARISON WITH EXACT EQUATIONS FOR SYSTEM AND SURROUNDINGS

In the preceding section, we assumed that $f(t)$ represented a set of given functions of space and time that are known from macroscopic observations. In this section, we obtain an expression for $f(t)$ from the exact equations of motion for the physical variables of the system and its surroundings.

Consider an isolated system consisting of the system of interest interacting purely dynamically with its surroundings. The total Hamiltonian is

$$\mathcal{H}(t)=\mathcal{H}_s(t)+\mathcal{H}_r(t)+\mathcal{H}_i, \quad (9)$$

and the operators F consist of two parts: F_s for the system of interest, and F_r for its surroundings. All system operators commute with all surroundings operators; however, no operators commute with the interaction $\mathcal{H}_i$ between the system and its surroundings. The generalized canonical density operator, which is defined[1] by Eq. (1–6), then becomes

$$\sigma=\sigma_s\sigma_r=\exp\,(-\lambda_s\cdot F_s-\lambda_r\cdot F_r)/Z_sZ_r, \quad (10)$$

where the conjugate functions λ are chosen so that Eq. (5) gives the correct expectations for the operators F. This can be used to express the initial condition (1)–(15) for the statistical density operator if we assume that $\mathcal{H}_i$ is zero for $t<0$ and that the statistical density operators of the system and of its surroundings are both initially generalized canonical density operators. The equations of motion (2–3) can now be separated into two exact coupled equations

$$\frac{\partial\langle F_s\rangle_t}{\partial t}=\langle\dot{F}_s\rangle_t+f_s(t)+\int_0^t dt'K_{s,s}(t,t')\cdot\lambda_s(t'), \quad (11)$$

and

$$\frac{\partial\langle F_r\rangle_t}{\partial t} = \langle \dot{F}_r\rangle_t + f_r'(t) + \int_0^t dt' K_{r,r}(t,t')\cdot\lambda_r(t'), \quad (12)$$

where $f_s'(t)$ and $f_r'(t)$ are defined by

$$f_s'(t) \equiv \int_0^t dt' K_{s,r}(t,t')\cdot\lambda_r(t') \quad (13)$$

and

$$f_r'(t) \equiv \int_0^t dt' K_{r,s}(t,t')\cdot\lambda_s(t'). \quad (14)$$

Here the subscripts on the kernels K refer to the subscripts on the operators F contained in them, as given in Eq. (2-4).

In the following we will show how Eq. (11) can be identified with Eq. (8), which was obtained under the assumption of thermal driving. Equation (13) then gives an expression for the thermal driving source in terms of the Hamiltonian (9) and the thermodynamic conjugate coordinate λ_r of the surroundings.

The terms in Eq. (11) will be identified one at a time. Notice that Eqs. (11) and (12) are coupled in all three terms on the right and not just by the f' terms.

In the $\langle \dot{F}_s\rangle_t$ term in Eq. (11) the coupling occurs only in the $\langle[\mathcal{H}_i, F_s]\rangle_t$ part since only this part depends upon λ_r as well as on λ_s. The remaining part does not couple the equations because $\langle[\mathcal{H}_s, F_s]\rangle_t$ depends only upon λ_s, and the rest is zero. The coupling in $\langle[\mathcal{H}_i, F_s]\rangle_t$ may be expressed as an average force exerted by the surroundings on the system. If, instead of considering the purely dynamical analysis of the system interacting with its surroundings, we consider the system as an isolated thermally and dynamically driven system, then this average force could be included as an external potential in the Hamiltonian of the system. The effective Hamiltonian for the system then is

$$\mathcal{H}_e(t) = \mathrm{Tr}_r\,[\mathcal{H}(t)\sigma_r], \quad (15)$$

which depends upon λ_r as it should. If this Hamiltonian is used in Eq. (8), then $\langle \dot{F}\rangle_t$ in that equation is exactly equal to the $\langle \dot{F}_s\rangle_t$ in Eq. (11). Hence $\langle \dot{F}_s\rangle_t$ in Eq. (11) can be identified with $\langle \dot{F}\rangle_t$ in Eq. (8), and the coupling in this term will occur only dynamically through the dependence of the average interaction on λ_r.

In the integral term in Eq. (11) the coupling occurs in the kernel $K_{s,s}$ since it depends upon λ_r as well as on λ_s. However, to a good approximation, the $\mathcal{H}$ used to compute this kernel can be replaced by $\mathcal{H}_e$. Then $K_{s,s}$ will depend upon λ_r only through its dependence on $\mathcal{H}_e$ and no longer through the σ used in the definition of K given in Eqs. (2-4)–(2-6), (2-9), and (2-10). Hence in this approximation $K_{s,s}$ in Eq. (11) can be identified with K in Eq. (8), and the coupling in this term will occur only dynamically through the dependence of the effective interaction on λ_r.

This completes the identification of Eq. (11) with Eq. (8). Therefore, the source strength $f'(t)$ is given by Eq. (13). This is a microscopic expression since the kernel $K_{s,r}$ is defined microscopically by Eq. (2-4). It shows that f' depends upon λ_r and also upon λ_s since $K_{s,r}$ depends upon both λ_s and λ_r. As a result, f' in Eq. (8) depends upon the thermodynamic state of both the system and its surroundings. Furthermore, even if both λ_s and λ_r are nearly constant in space, f' will have a spatial dependence determined by the kernel $K_{s,r}$ in Eq. (13). This leads to a limitation on the kind of problem to which the thermal driving formalism can be applied.

The idea behind the thermal driving formalism is that the source strength f' is known experimentally and need not be calculated microscopically as in this section. Hence a limitation to the usefulness of the thermal driving formalism arises because experimental knowledge of the space and time dependence of f' is limited in practice. Fine details are not observed except with great effort, and, if these details are important, the formalism is not useful.

For example, consider the calculation of the Kapitza resistance to heat flow across a surface between two dissimilar materials. We might try to compute the spatial dependence of the temperature near the surface by considering one material as the system, with heat entering from outside. Now although the total energy flux into the system is easily measured, we do not know its precise spatial dependence. If we assume that the entire source is concentrated on the surface, then we find an infinite jump in the temperature at the surface. This can be seen with a simple steady-state calculation assuming an exponential spatial dependence for the kernel. In order to avoid the infinite temperature jump, we must distribute the source near the surface in a region whose thickness is of the order of the correlation length in the energy-current autocorrelation function (2-27). But we cannot use an arbitrary distribution since the spatial dependence of the temperature near the surface depends on it. The distribution can be determined only by considering both materials together as an isolated system and using the exact formalism of this section or of Refs. 1 and 2. But then the equations of motion can be solved directly for the spatial dependence of the temperature near the surface of interest, and there is no need to consider how the heat goes from one material to the other.

This example brings out a general limitation to the usefulness of the thermal driving formalism. This formalism is not useful for determining the detailed response of a system within approximately a correlation length from where the system is thermally driven. However, in practice, this limitation is often not relevant since the detailed response within a correlation length is not of interest. This is shown empirically by the wide applicability and success of the equations of nonequilibrium thermodynamics, which we have derived from our definition of thermal driving in statistical mechanics.

ACKNOWLEDGMENT

One of us (W. C. M.) would like to thank Professor E. T. Jaynes for suggesting the problem of formulating transport into a system in statistical mechanics.

* NRC–NBS Postdoctoral Research Associate, 1969–71.

[1] B. Robertson, Phys. Rev. **144**, 151 (1966). Equation (n) of this reference will be denoted in the present paper by Eq. (1–n).

[2] B. Robertson, Phys. Rev. **160**, 175 (1967); **166**, 206 (1968), Erratum. Equation (n) of this reference will be denoted in the present paper by Eq. (2–n).

[3] B. Robertson, Phys. Rev. **153**, 391 (1967).

[4] B. Robertson, J. Math. Phys. **11**, 2482 (1970).

[5] Thermal driving was first defined in a slightly different way in a dissertation (unpublished) submitted by W. C. M. in 1967 to Washington University, St. Louis. The present paper gives a clearer exposition of the essential features of that work (Footnote 14) and connects it with Refs. 1 and 2.

[6] J. A. McLennan, Advan. Chem. Phys. **5**, 261 (1963).

[7] D. N. Zubarev, Dokl. Akad. Nauk SSSR **164**, 537 (1965) [Sov. Phys. Dokl. **10**, 850 (1966)].

[8] P. G. Bergmann and J. L. Lebowitz, Phys. Rev. **99**, 578 (1955); J. L. Lebowitz and P. G. Bergmann, Ann. Phys. (N.Y.) **1**, 1 (1957).

[9] G. G. Emch and G. L. Sewell, J. Math. Phys. **9**, 946 (1968).

[10] Equations (2–B2) and (2–A4).

[11] For most sets of operators $F_1(\mathbf{r}), \cdots, F_m(\mathbf{r})$, this bar is defined by Eq. (2–6). However, if there exists a ζ satisfying $\zeta \cdot F = 1$, then, as pointed out in the Erratum for Ref. 2, Eqs. (2–C2)–(2–C4) must be used in place of Eqs. (2–2), (2–6), and (2–7). In this case, these equations are also necessary in order to make $\langle A\bar{A}\rangle > 0$ for all A so that $\langle A\bar{B}\rangle$ can be an inner product of the operators A and B.

[12] Since $\langle A\bar{B}\rangle$ is an inner product (Footnote 11), a necessary and sufficient condition for this inverse to exist is that the operators $F_1(\mathbf{r}), \cdots, F_m(\mathbf{r})$ be linearly independent. See, e.g., N. I. Akhiezer and I. M. Glazman, *Theory of Linear Operators in Hilbert Space* (Ungar, New York, 1961), p. 12.

[13] To preserve normalization, the trace of the right side of Eq. (7) must be zero. Usually this will automatically be true because of Eqs. (2–5) and (2–6). However, if there exists a ζ satisfying $\zeta \cdot F = 1$, then (Footnote 11) Eqs. (2–C2)–(2–C4) must be used in place of Eqs. (2–2), (2–6), and (2–7). In this case the trace of the right side of Eq. (7) will no longer be automatically zero. In order to make it zero, $\dot{f}(t)$ must be restricted to satisfy $\zeta \cdot \dot{f}(t) = 0$. This restriction applies only if there exists a ζ satisfying $\zeta \cdot F = 1$, and then it is required in order to conserve probability.

[14] If instead we choose the $\sigma(t)$ in Eqs. (5) and (7) above to be just $\rho(t)$, then a formal solution to Eq. (7) is

$$\rho(t) = U(t) \exp\left(-\beta \mathcal{H}(0) + \int_0^t dt' \lambda(t') \cdot U^{-1}(t') F U(t')\right) U^{-1}(t),$$

where

$$i\hbar \frac{\partial U(t)}{\partial t} = \mathcal{H}(t) U(t), \quad U(0) \equiv 1,$$

and

$$\lambda(t) \equiv \dot{f}(t) \cdot \langle F\bar{F}\rangle_t^{-1}.$$

These equations were first derived in the dissertation cited in Ref. 5. There they were used to obtain Green–Kubo expressions for thermal transport coefficients in the necessary limits but without further assumption. However, they do not permit a simple generalization of the equation of motion for $\langle F\rangle_t$ when the system is driven thermally. Because of this and because of the many advantages of the equation of motion approach, the equations of this footnote are not used in this paper.

[15] In order to derive Eq. (8), insert Eq. (7) into Eq. (1–22) to get

$$\frac{\partial \sigma(t)}{\partial t} + P(t) \frac{i}{\hbar} [\mathcal{H}(t), \rho(t)] = \bar{F}\sigma(t) \cdot \langle F\bar{F}\rangle_t^{-1} \cdot \dot{f}(t).$$

When this equation is subtracted from Eq. (7), the source terms cancel, and the result is an equation of motion for $\rho(t)$–$\sigma(t)$ identical to the one obtained in Ref. 1. Hence, the integrating factor $T(t, t')$ for this equation is the same as before, and the solution

$$\rho(t) = \sigma(t) - \int_0^t dt' T(t, t')[1 - P(t')] iL(t') \sigma(t')$$

is also the same as before. This nonequilibrium statistical density operator is valid for systems arbitrarily far from equilibrium, either driven dynamically through $\mathcal{H}(t)$, or driven thermally through $\dot{f}(t)$. The thermal driving has been incorporated into the formalism even though $\dot{f}(t)$ does not appear explicitly in $\sigma(t)$, $P(t)$, $T(t, t')$, or this last equation; this is true because these operators depend upon $\langle F\rangle$, which itself depends upon $\dot{f}$, as can be seen in Eq. (8).

The equation of motion for $\langle F\rangle_t$ is obtained, as before, by inserting the last equation into Eq. (7), multiplying by F, and taking the trace. When the steps leading to Eq. (1–32) are taken, the result is Eq. (8) above.

[16] The kernel $K(t, t')$ is defined in Eq. (2–4), and the conjugate function $\lambda(t)$ is defined in Eq. (2–2).